



Escola d'Enginyeria de Telecomunicació i  
Aeroespacial de Castelldefels

UNIVERSITAT POLITÈCNICA DE CATALUNYA

# DETECCIO DEL XIULET PER LA GENERACIO DE RESUMS EN VIDEOS ESPORTIUS

**TÍTOL DEL PFC:** Detecció del xiulet per la generació de resums en vídeos esportius

**TITULACIÓ:** Enginyeria de Telecomunicació (segon cicle)

**AUTOR:** Sergi Sicília Sansi

**DIRECTOR:** Francesc Tarrés Ruiz

**DATA:** 14 de febrer de 2013



**Títol:** Detecció del xiulet per la generació de resums en vídeos esportius

**Autor:** Sergi Sicília Sansi

**Director:** Francesc Tarrés Ruiz

**Data:** 14 de febrer de 2013

## **Resum**

El present document exposa el desenvolupament de l'aplicació del mètode d'Anàlisi de Components Principals per detectar el xiulet de l'àrbitre en el context de l'extracció automàtica d'esdeveniments d'interès en vídeos esportius.

L'extracció automàtica d'esdeveniments d'interès de vídeos esportius consisteix en un sistema capaç de detectar en un vídeo esportiu els esdeveniments de major interès, per tal d'extraure'ls i fer-ne un resum en el qual quedin etiquetades les jugades d'interès de tal manera que en quedi facilitat el posterior accés.

El sistema empra l'anàlisi d'àudio i vídeo a partir de descriptors de baix nivell i altres elements, com ara la detecció del xiulet de l'àrbitre el qual constitueix l'objecte principal del present document.

Al llarg del document es tenen en compte les propietats tècniques dels xiulets més emprats en l'arbitratge europeu de futbol a partir de la seva caracterització. La dificultat tècnica que presenta la detecció del xiulet queda evidenciada pels primers mètodes emprats. Finalment es presenta el mètode emprat en la solució final.

La solució està basada en l'aplicació del mètode d'Anàlisi de Components Principals (PCA, Principal Component Analysis), mètode conegut per la seva aplicació en la detecció d'objectes complexos i que es basa en un procediment matemàtic que transforma un nombre de variables correlacionades en un nombre més petit de variables no correlacionades anomenades components principals. Aquest procediment permet descobrir l'estructura comuna del conjunt de dades i emprar-la per a la detecció de similituds en nous conjunts de dades.

Els resultats de l'aplicació del mètode de PCA són satisfactoris però es discuteix, tanmateix, en quin grau compleixen amb la precisió requerida en el camp de l'extracció automàtica d'esdeveniments d'interès en vídeos esportius. El document presenta línies futures de treball per procurar millorar la precisió i conclou exposant la possibilitat d'aplicar el PCA en la detecció de xiulets donats els positius resultats en la seva exhaustivitat i exactitud.

**Title:** Whistle Detection for summary sport videos generation

**Author:** Sergi Sicília Sansi

**Director:** Francesc Tarrés Ruiz

**Date:** February, 14th 2013

## Overview

This report presents the development of an algorithm that uses the Principal Component Analysis method to detect the referee's whistle in sport videos. The detection of the referee's whistle is very important in the context of automatic interest event extraction in sport videos due to the fact that many events of major interest have as a result or cause a referee whistle sound.

Automated interest event extraction in sport videos consists in a system capable of detecting events of major interest in sport videos, in order to extract them and make a summary in which the interesting events are labelled and easily indexed.

The system uses the analysis of low-level audio and video descriptors and other elements, such as the detection of the referee's whistle, which constitutes the main object of this document.

This report takes into account the properties of the most used whistles in the arbitration of European football. These properties are taken from the characterization of two whistle models. The technical difficulty to detect the whistle sound is shown by the first methods used. Finally, the method used in this work is presented in the final solution.

The final solution is based on the application of the method of Principal Component Analysis (PCA), method known for its application in the detection of complex objects. It is based on a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables called principal components. This procedure reveals the common structure of the data set and uses it to detect similarities in new data sets.

The results of applying the PCA method are satisfactory and the report discusses, however, in which extent they meet the required accuracy in the field of interest event automatic extraction in sport videos. The paper presents future lines to try to improve the precision and concludes by discussing on the use of the PCA method in the detection of referee whistles given the positive results as far as recall and accuracy are concerned.

# ÍNDEX

<b>INTRODUCCIÓ .....</b>	<b>1</b>
<b>CAPÍTOL 1. DETECCIÓ DEL XIULET EN L'EXTRACCIÓ AUTOMÀTICA DE VÍDEOS ESPORTIUS .....</b>	<b>3</b>
1.1. Extracció automàtica d'esdeveniments d'interès de vídeos esportius .....	3
1.2. Detecció automàtica de xiulets d'àrbitre.....	5
<b>CAPÍTOL 2. CARACTERITZACIÓ DELS XIULETS I MÈTODES DE DETECCIÓ.....</b>	<b>8</b>
2.1. Història del xiulet en l'ús d'arbitratge esportiu .....	8
2.2. Tipus de xiulets emprats en l'arbitratge futbolístic .....	10
2.3. Caracterització dels xiulets .....	12
2.3.1. Fox 40.....	12
2.3.2. Naysir N1B.....	14
2.4. Primers mètodes emprats per la detecció de xiulets d'àrbitre .....	17
<b>CAPÍTOL 3. ANÀLISI DE COMPONENTS PRINCIPALS (PCA).....</b>	<b>21</b>
3.1. Descripció del mètode d'Anàlisi de Components Principals.....	21
3.2. Procediments del PCA .....	22
<b>CAPÍTOL 4. PCA APLICAT A LA DETECCIÓ DE XIULETS .....</b>	<b>25</b>
4.1. Diagrama de blocs.....	25
4.1.1. Compilació de DFTs de xiulets mostra.....	27
4.1.2. Càlcul de la matriu de covariança .....	28
4.1.3. Càlcul dels autovectors i autovalors .....	28
4.1.4. Entrada de la pista d'àudio .....	29
4.1.5. Càlcul de la DFT del segment .....	30
4.1.6. Càlcul de les autocoordenades .....	31
4.1.7. Reconstrucció .....	32
4.1.8. Criteri de decisió .....	37
4.1.9. Post-processat.....	38
<b>CAPÍTOL 5. PRESENTACIÓ DE RESULTATS .....</b>	<b>39</b>
5.1. Precisió, exactitud i exhaustivitat.....	39
5.2. Resultats.....	41
<b>CAPÍTOL 6. LÍNIES FUTURES DE TREBALL .....</b>	<b>44</b>

<b>CONCLUSIONS.....</b>	<b>46</b>
<b>BIBLIOGRAFIA CONSULTADA .....</b>	<b>49</b>

# INTRODUCCIÓ

El present document presenta una proposta de solució a la detecció del xiulet de l'àrbitre en el context de l'extracció automàtica d'esdeveniments d'interès de vídeos esportius, concretament de vídeos de futbol de la Liga Profesional de Fútbol espanyola.

L'extracció automàtica d'esdeveniments d'interès de vídeos esportius consisteix en un sistema capaç de detectar en un vídeo esportiu els esdeveniments de major interès, per tal d'extraure'ls i fer-ne un resum en el qual quedin etiquetades les jugades d'interès de tal manera que en quedi facilitat el posterior accés.

Per tal d'aconseguir identificar les jugades d'interès s'empren anàlisis d'àudio i vídeo a partir de descriptors de baix nivell i altres elements, com és la detecció del xiulet de l'àrbitre. És precisament aquest últim element el que es presenta com a objectiu del sistema que descriu el present document.

La solució està basada en el mètode d'Anàlisi de Components Principals (PCA, Principal Component Analysis), mètode conegut per la seva aplicació en la detecció d'objectes complexos i que es basa en un procediment matemàtic que transforma un nombre de variables correlacionades en un nombre més petit de variables no correlacionades anomenades components principals. Aquest procediment permet descobrir l'estructura comuna del conjunt de dades i aplicar-la per a la detecció de similituds en nous conjunts de dades.

Per tal de mostrar les raons de l'aplicació del mètode d'Anàlisi de Components Principals per a la detecció de xiulets de l'àrbitre i els resultats de dita aplicació, el document es divideix en sis capítols.

El primer capítol exposa en què consisteix l'extracció automàtica d'esdeveniments d'interès en vídeos esportius i en quina mesura la detecció del xiulet de l'àrbitre n'és un dels components principals donada la seva relació amb la presència de jugades d'interès.

El capítol segon procedeix a presentar la història del xiulet en l'arbitratge esportiu i els tipus de xiulet que s'empren per arbitrar els partits de futbol. Així mateix, el segon capítol presenta una caracterització dels dos xiulets més emprats, un model de xiulet que empra una bola de cautxú en la cavitat interior i un model de xiulet que no empra bola sinó que emet un so compost de tres pics freqüencials. Al final d'aquest capítol també es presenten els primer mètodes emprats per a intentar detectar el xiulet de l'àrbitre en partits de futbol de la Liga Profesional de Fútbol espanyola.

A partir de l'explicitació de les característiques tècniques dels xiulets i les dificultats trobades en els primers intents de resoldre el problema, el capítol tercer presenta el mètode emprat, conegut com Anàlisi de Components

Principals (PCA). Se n'explicita la seva descripció i es para atenció als seus procediments més importants.

El capítol quart constitueix el capítol més llarg del document donat que és el que presenta el desenvolupament de l'Anàlisi de Components Principals en la seva aplicació per a la detecció del xiulet de l'àrbitre. Aquest capítol descriu el diagrama de blocs del sistema i cadascuna de les seves etapes implementades amb el programa Matlab de Mathworks.

Després d'haver-ne explicat el funcionament, el capítol cinquè presenta els resultats obtinguts. Els resultats corresponen a l'aplicació del sistema a partits de futbol enregistrats de la LPF i s'expressen a partir dels termes matemàtics de precisió, exhaustivitat i exactitud. Com es veurà, els resultats mostren una alta exactitud i exhaustivitat però a canvi d'una precisió inferior. En les conclusions es discuteix en quina mesura afecta això en la implementació de la detecció de xiulets esportius en l'extracció automàtica d'esdeveniments d'interès en vídeos esportius.

El capítol sisè presenta les línies futures de treball per millorar el rendiment del sistema a partir dels resultats obtinguts.

El document finalitza amb les conclusions en les quals es discuteix la capacitat del mètode d'Anàlisi de Components Principals per detectar el xiulet de l'àrbitre en l'àmbit de l'extracció automàtica d'esdeveniments d'interès en vídeos esportius. En resum, se suggereix que el sistema precisa ser millorat en algunes de les seves parts per tal d'aconseguir resultats més favorables i, per tant, no es descarta el mètode del PCA en l'aplicació de la detecció de xiulets de l'àrbitre.



## **CAPÍTOL 1. DETECCIÓ DEL XIULET EN L'EXTRACCIÓ AUTOMÀTICA DE VÍDEOS ESPORTIUS**

Al llarg d'aquest primer capítol s'explica el context general del què suposa l'extracció automàtica d'esdeveniments d'interès en vídeos esportius i l'especificitat del present projecte que consisteix en detectar el xiulet de l'àrbitre. Tal i com es veurà a continuació, la detecció del xiulet de l'àrbitre és un dels elements més importants per ajudar a la detecció automàtica d'esdeveniments d'interès en vídeos esportius ja que aquesta tècnica emprava diversos descriptors i elements d'àudio i vídeo per identificar els esdeveniments que es pretenen extreure automàticament.

Els descriptors que s'empren s'extreuen de l'anàlisi separat de l'àudio i el vídeo que s'està analitzant. A partir d'aquest anàlisi es pot efectuar un creuament de la informació obtinguda per detectar de quin tipus d'esdeveniment d'interès es tracta. Així, un esdeveniment concret consistirà en una combinació de descriptors de tal manera que els esdeveniments d'importància s'aniran classificant i etiquetant automàticament en la mesura en què es realitza l'anàlisi del vídeo esportiu.

La raó per la qual la detecció del xiulet de l'àrbitre és tan important queda palesa també en aquest capítol. Els àrbitres empren el xiulet per indicar els moments en els quals s'ha d'aturar o reiniciar el joc, seguint unes normes que dicta la FIFA, o bé en qualsevol moment que es necessita cridar l'atenció dels jugadors. D'aquesta manera moltes de les jugades d'interès, tot i que no necessàriament totes, tindran com a inici o com a conseqüència un xiulet de l'àrbitre.

### **1.1. Extracció automàtica d'esdeveniments d'interès de vídeos esportius**

La quantitat d'informació esportiva enregistrada augmenta dia rere dia fet possible, en part, per la creixent capacitat d'emmagatzematge digital. És per això que l'extracció automàtica d'esdeveniments d'interès es fa cada dia una mica més preuada. Aquesta consisteix en un sistema capaç de detectar en un vídeo esportiu els esdeveniments de major interès, per tal d'extraure'ls i fer-ne un resum en el qual quedin etiquetades les jugades d'interès de tal manera que en quedi facilitat el posterior accés.

L'extracció automàtica d'esdeveniments d'interès de vídeos esportius [1] s'acostuma a realitzar a partir d'un processat digital de la senyal del vídeo separatament de la de l'àudio, intentant localitzar els elements que constitueixen una característica especial del joc i que per tant han de ser identificats i seleccionats si es vol elaborar un resum automàtic d'aquest.

En primer terme, la informació de vídeo sol ser molt important ja que durant la transmissió de l'esdeveniment esportiu el realitzador utilitza una sèrie de recursos d'estil que poden indicar la presència de jugades d'interès. Entre aquests recursos es poden incloure les repeticions de jugades, els zooms, la captura de jugadors en primers plans, els moviments ràpids de càmera, així com altres recursos possibles.

En segon terme, la informació proporcionada per l'anàlisi del so ambient també presenta una gran importància ja que, generalment, les reaccions del públic mostren una relació directament proporcional amb l'interès que puguin tenir les jugades. És per aquesta raó que poden emprar-se descriptors de molt baix nivell com la potència del so o l'evolució de l'espectre i relacionar-los amb el possible interès d'una jugada. De tots els elements que es poden analitzar de l'àudio, un dels més importants és el xiulet de l'àrbitre, ja que aquest es pot relacionar amb un tipus concret de jugades d'interès, ja que l'ús del xiulet està previst pel reglament i per tant, generalment, el xiulet de l'àrbitre indicarà un nombre concret de possibles fets.

Tal i com s'acaba d'explicar, la segona fase de l'extracció automàtica d'esdeveniments d'interès se centra en la detecció i classificació [2] de dits esdeveniments a partir de la informació extreta de l'anàlisi de la primera etapa. Aquest procés és dut a terme a través d'un conjunt de filtres que s'activen per certes combinacions de descriptors de baix nivell.

Els descriptors d'àudio de baix nivell [3] representen la variació de les propietats d'àudio ja sigui en el domini temporal o freqüencial. Gràcies a aquests descriptors és possible analitzar la similitud entre diferents arxius d'àudio de manera que és possible identificar àudio de contingut idèntic, semblant o diferent.

Així, emprant els descriptors de baix nivell és possible definir un esdeveniment d'interès com una certa combinació de descriptors. Alguns exemples d'esdeveniments d'interès en un partit de futbol són les faltes, gols, repeticions, celebracions, córners, fores de joc, jugades perilloses, etc.

El resultat final es presenta com un resum automàtic del partit que conté aquest conjunt d'esdeveniments d'interès detectats i que descriuen el contingut audiovisual analitzat.

Per tal de poder realitzar una detecció automàtica dels esdeveniments d'interès és necessari realitzar un anàlisi previ basat en l'ocurrència dels descriptors que s'empren per detectar dits esdeveniments. Per exemple, un gol podria definir-se com una combinació formada per una pujada d'àudio per part del públic, la presència de persones ocasionada per la celebració dels jugadors, la detecció d'una repetició de la jugada i la d'un xiulet conforme es reprèn el joc al final de l'esdeveniment.

Així, podríem definir l'objectiu dels sistemes d'extracció automàtica d'esdeveniments d'interès de vídeos esportius com l'agrupació lògica de

l'ocurrència de diversos descriptors de baix nivell per construir esdeveniments amb algun significat semàntic.

## 1.2. Detecció automàtica de xiulets d'àrbitre

La detecció automàtica de xiulets d'àrbitre [1] és una tasca important donat que moltes jugades d'interès presenten un xiulet de l'àrbitre com a inici o com a conseqüència. No es tracta només del cas del futbol sinó que, en general, tots els esports reglats empen de la persona d'un jutge o àrbitre per garantir que es compleixin les regles del joc i determinar quan un equip o jugador assoleix els objectius del joc, ja siguin punts o gols. Igualment, el jutge o àrbitre s'encarrega d'establir les possibles penalitzacions quan un dels jugadors no compleix amb les normes que preveu la regulació de l'esport que s'està practicant.

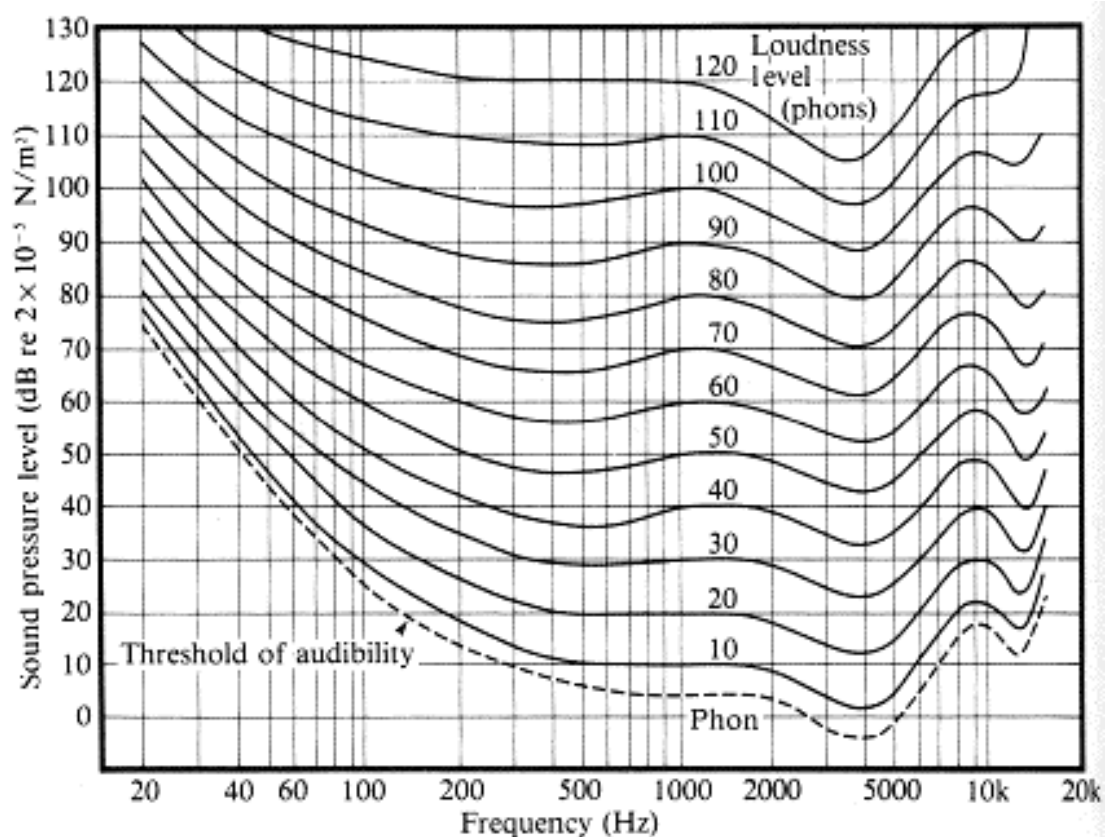
L'ús d'imatges capturades per diferents càmeres per realitzar un anàlisi detallat de les circumstàncies del joc només s'admet en alguns esports. En alguns casos, certes decisions arbitrals poden ser preses un cop s'ha finalitzat la competició. És el cas de la Fórmula 1 o l'atletisme.

Però no tots els esports contempnen la possibilitat que els àrbitres emprin aquestes tècniques un cop s'ha acabat la competició sinó que es prefereix que les decisions es prenguin al moment. Així, trobem el cas, per exemple, del futbol americà, que tot i emprar imatges capturades per càmeres, les decisions són preses en el transcurs del partit. En canvi, en el futbol europeu s'ha preferit que aquestes decisions instantànies dels àrbitres no puguin fer-se amb l'ajuda d'imatges gravades.

En els esports on la decisió de l'àrbitre ha de ser instantània i el soroll ambient produït pel públic tendeix a ser gran, s'acostuma a emprar l'ús d'un xiulet perquè l'àrbitre senyalitzi de forma acústica en quin moment s'ha d'aturar o reprendre el joc.

El xiulets estan pensats per produir una senyal acústica que es pugui sentir o bé en llocs on el soroll ambient és molt gran o bé a llargues distàncies. És per això que produeixen un so de gran potència. Alguns, com és el cas del Fox 40, s'acosten al marge superior dinàmic de l'oïda humana [4] el qual se situa entre els 0 dB ( $2 \cdot 10^{-5}$  Pascals) i els 120 dB (20 Pascals), arribant a zones properes al llindar de dolor auditiu. A més, generalment, la zona de freqüències en la qual es troba el pic o pics de potència del xiulet acostuma a ser una zona especialment sensible per a l'oïda humana.

La Fig. 1.1 mostra les corbes [5] d'igual volum acústic per l'oïda humana, en les quals es pot observar com a les freqüències entorn els 4 kHz l'oïda humana és especialment sensible sobretot per les corbes que corresponen a major volum acústic.



**Fig. 1.1** Corbes de Fletcher-Munson, corbes d'igual potència sonora en funció de la freqüència

Tanmateix, avui en dia, donades les grans dimensions dels estadis i la gran quantitat d'espectadors que poden assistir als recintes esportius, no sempre és fàcil detectar la presència del xiulet de l'àrbitre. Encara que, sense cap dubte, la tecnologia actual permetria usar un altre tipus d'elements de senyalització molt més fiables que el xiulet, aquest se segueix utilitzant, per tradició, en la major part d'esports.

En el cas concret del futbol, el reglament de la FIFA [6] (Fédération Internationale de Football Association) estableix les circumstàncies del joc en què l'àrbitre ha d'usar el xiulet:

- Serà necessari emprar el xiulet en les següents situacions:
  - Per iniciar el joc a la primera part, la segona part i després d'un gol.
  - Per aturar el joc a causa de:
    - Concedir un tir lliure o un tir penal.
    - Suspendre o finalitzar un partit.
    - Finalitzar un període del joc perquè el temps ha expirat.
  - Per reiniciar el joc en:

- Tirs lliures, al ordenar que la barrera se situï a la distància apropiada.
  - Tirs penals.
- Per reiniciar el joc si ha estat detingut per:
  - Aplicació d'una targeta groga o vermella per una incorrecció.
  - Lesió,
  - Substitució.
- És opcional l'ús del xiulet per:
  - Interrompre el joc per:
    - Un servei de porta
    - Un servei de córner
    - Un servei de banda.
    - Un gol
  - Reiniciar el joc després de:
    - Un tir lliure
    - Un servei de porta
    - Un servei de córner
    - Un servei de banda

La normativa també estableix que quan es vulgui emprar el xiulet per reiniciar el joc, l'àrbitre anunciarà clarament als jugadors que el joc no es reprendrà fins després de dita senyal. Per altra banda, el reglament també adverteix que el xiulet que s'empri innecessàriament amb massa freqüència tindrà menys impacte a la llarga.

D'acord amb aquesta normativa, podem concloure que detectar amb fiabilitat el xiulet de l'àrbitre és un pas important per localitzar amb precisió un bon nombre d'esdeveniments d'interès com llançaments de penals, llançaments de faltes directes amb barrera, inici i final de joc, etc.

El reglament no garanteix que es detecti tot tipus de jugades pel fet que en molts casos es deixa a criteri de l'àrbitre la senyalització acústica dels esdeveniments (gol, córner, etc.). Tanmateix, encara que la detecció del xiulet no proporcioni localització de tots els esdeveniments d'interès del partit sí que cadascun dels xiulets es correspon amb esdeveniments d'interès que poden haver ocorregut abans o després d'aquest.

Per tant, la detecció del xiulet de l'àrbitre és una de les etapes més importants en l'extracció automàtica d'esdeveniments d'interès en vídeos esportius donat que la major part de les jugades d'interès tenen com a resultat o origen un xiulet de l'àrbitre.

## **CAPÍTOL 2. CARACTERITZACIÓ DELS XIULETS I MÈTODES DE DETECCIÓ**

Un cop s'ha vist la importància de la detecció del xiulet de l'àrbitre per a l'extracció automàtica d'esdeveniments d'interès en vídeos esportius, el segon capítol es proposa desenvolupar l'estudi de les característiques tècniques dels xiulets emprats en l'arbitratge per tal de justificar-ne el mètode emprat per a la seva detecció.

El xiulet no s'havia emprat sempre en l'arbitratge esportiu. De fet, es creu que es va emprar per primera vegada el 1878. La raó per la qual es va emprar per primera vegada i la raó per la qual després d'aquest primer ús es va continuar emprant no és indiferent, sinó que constitueix una de les característiques tècniques del xiulet: es va emprar per cridar l'atenció dels jugadors i va tenir un resultat molt eficient. Certament, el so d'un xiulet es produeix en una zona freqüencial de gran sensibilitat auditiva humana i per això aconseguí despertar la resposta dels jugadors amb tanta eficàcia.

Si bé qualsevol xiulet acostuma a generar un so del tipus descrit, la varietat de xiulets presents en el mercat és molt àmplia. En aquest capítol es caracteritzen dos xiulets de famílies diferents que són els més emprats en l'arbitratge esportiu.

En primer lloc, i després d'explicar-ne el seu origen històric, es caracteritza el xiulet Fox 40 el qual és un xiulet que no empra bola de cautxú en la cavitat interior sinó que realitza un so amb tres pics de potència de diferent freqüència. En segon lloc, es caracteritza el Naysir N1B el qual és un dels xiulets més populars que empra bola de cautxú en la cavitat interior.

Finalment, el capítol acaba exposant, com a conclusió de la caracterització tècnica dels esmentats xiulets, diferents tècniques que es van emprar inicialment per tal d'aconseguir una eficaç detecció del xiulet de l'àrbitre en vídeos esportius i com aquestes tècniques no van donar resultat raó per la qual es va haver d'emprar una alternativa, que en el cas present es tracta de l'Anàlisi de Components Principals (en anglès Principal Component Analysis, PCA).

### **2.1. Història del xiulet en l'ús d'arbitratge esportiu**

L'ús del xiulet [7] en la història humana ha pres protagonisme en diferents àmbits al llarg del temps. Fora de l'àmbit esportiu, els xiulets d'os o de fusta s'han utilitzat durant milers d'anys. Així, s'han emprat al llarg de la història en els vaixells de guerra per emetre ordres i salutacions, en l'antiga Grècia i Roma eren emprats per seguir les carreres de galeres d'esclaus i una versió medieval va ser utilitzada durant les Croades per reunir els anglesos a coberta davant d'un imminent atac.

L'era moderna de l'ús del xiulet va començar el 1878 quan un xiulet va ser emprat per primera vegada per un àrbitre durant un esdeveniment esportiu. Hudson, un fabricant d'eines que estava fascinat amb els xiulets, va fer un instrument de metall que es va utilitzar en un partit en el club de futbol Nottingham Forest. L'ús del xiulet va donar millor resultat que el senyal de l'àrbitre usual d'agitar un mocador cosa que va fer que es pensés en la seva utilitat en el món de l'esport.

A partir d'aquell moment, i veient l'eficàcia que havia obtingut l'ús del xiulet per captar l'atenció dels jugadors, les normatives i els reglaments de diferents esports d'equip que es juguen amb públic abundant començaren a incorporar el xiulet com a eina d'arbitratge a favor del jutge. En alguns esports, com és el cas del bàsquet, es combinen xiulets convencionals (àrbitre) amb xiulets electrònics (jutge de taula) mentre que en altres esports només s'utilitzen els xiulets convencionals.

El 1883, el cos de policia de Londres va donar a conèixer que estava buscant un objecte que produís un so eficaç per tal que els agents poguessin fer-lo servir per cridar l'atenció de la gent. Hudson va inventar un xiulet lleuger i compacte que produïa dos tons discordants audibles a més de 1,5 km. El xiulet va ser adoptat immediatament i el mateix disseny encara està en ús avui en dia.

L'any següent, Hudson va inventar el "xiulet de bola" [1]. Aquest seria, amb el temps, el tipus de xiulet més conegut pel públic en general. Consta d'una petita cavitat cilíndrica, amb un orifici en un dels seus costats, pel qual circula l'aire a gran velocitat, produint un so d'una freqüència elevada quan l'aire passa a través de l'orifici. A l'interior de la cavitat hi ha una petita bola de cautxú que es mou i que eventualment dificulta el flux de l'aire a través de l'orifici interrompent breument la producció del so. Aquestes interrupcions breus i ràpides del flux d'aire caracteritzen el so del xiulet i fan que sigui adequat per cridar l'atenció. A la Fig. 2. es mostra un típic xiulet de bola.



**Fig. 2.1** Exemple de xiulet de bola.

El problema principal dels xiulets de bola és que en determinades circumstàncies, si el nivell de flux de l'aire és molt elevat, és possible que la bola quedi atrapada temporalment a l'orifici i que el xiulet deixi de sonar. La correcta vibració de la bola també pot variar amb la humitat, les condicions atmosfèriques (pluja), etc. En definitiva, es tracta d'un tipus de xiulet poc precís raó per la qual és poc usat en l'arbitratge actual.

Com que els xiulets de bola presentaven tots aquests inconvenients Ron Foxcroft [8], que era àrbitre de bàsquet, va inventar un nou disseny basat en els instruments de vent harmònicament afinats. La diferència fonamental era que el nou xiulet produeix tres freqüències bastant pròximes, que en interferir entre si produïen un so molt semblant al d'un xiulet de bola convencional.

El primer model fabricat per Ron Foxcroft es va denominar Fox 40 i ha estat utilitzat com a xiulet oficial en nombroses competicions oficials de diferents esports. A part del Fox 40 hi ha altres fabricants i models que poden estar recomanats en algunes competicions. En alguns d'aquests models les freqüències dels tons poden canviar sensiblement i fins i tot, en alguns casos, utilitzar xiulets bitonals, en comptes de tritonals com és el cas del xiulet Fox40.

Les característiques del so produït per un xiulet [1] són molt especials i produeixen una reacció immediata en el cervell humà. Per això, el seu ús no es limita a esdeveniments esportius sinó que també és usat per la policia per cridar l'atenció, regular el trànsit, etc, en determinats tipus de música, com la samba, com a instrument rítmic no percussiu i també és emprat per situacions d'emergència i rescat en activitats a la naturalesa.

Encara que en el següent apartat veurem amb detall les seves característiques és important esmentar que les freqüències que produeix estan situades a la zona de major sensibilitat del sistema auditiu humà.

A més, la ràpida variació de l'amplitud del senyal generada, ja sigui produïda per la bola o per la interferència entre freqüències, produeix un so molest que captura l'atenció del cervell.

Finalment, la potència de so produïda per un xiulet professional pot situar-se al voltant dels 122 dB, pròxims al llindar de dolor del sistema auditiu humà. Aquests tres factors: freqüències, vibració i potència fan que segueixi usant-se el xiulet en molts esports d'equip on el nivell de soroll ambient produït pel públic pot arribar a ser considerable.

## **2.2. Tipus de xiulets emprats en l'arbitratge futbolístic**

El tipus de xiulet emprat acostuma a dependre de la competició en la qual s'hagi de dur a terme l'arbitratge. Generalment l'ús de xiulets sense bola en la cavitat inferior és preferit perquè presenta major fiabilitat i eficàcia. A Europa, l'ús del xiulet model Fox 40 és el més estès i de fet és el xiulet oficial en moltes competicions de futbol europeu. El segon xiulet més emprat que no forma part



de la família dels Fox 40 és el Naysir N1B. Aquest seran els dos xiulets que s'analitzaran a continuació i que en el següent apartat seran caracteritzats.

Les característiques tècniques del xiulet Fox 40 són les següents [9]:

- Potència sonora: 115 dB.
- Disseny en 3 cambres sense bola a la cavitat interior.
- So compostat de 3 freqüències principals.



**Fig. 2.2** Xiulet model Fox 40.

Les característiques tècniques del xiulet Naysir N1B són les següents:

- So al voltant de la freqüència central de 2900 Hz.



**Fig. 2.3** Xiulet model Naysir N1B.

## 2.3. Caracterització dels xiulets

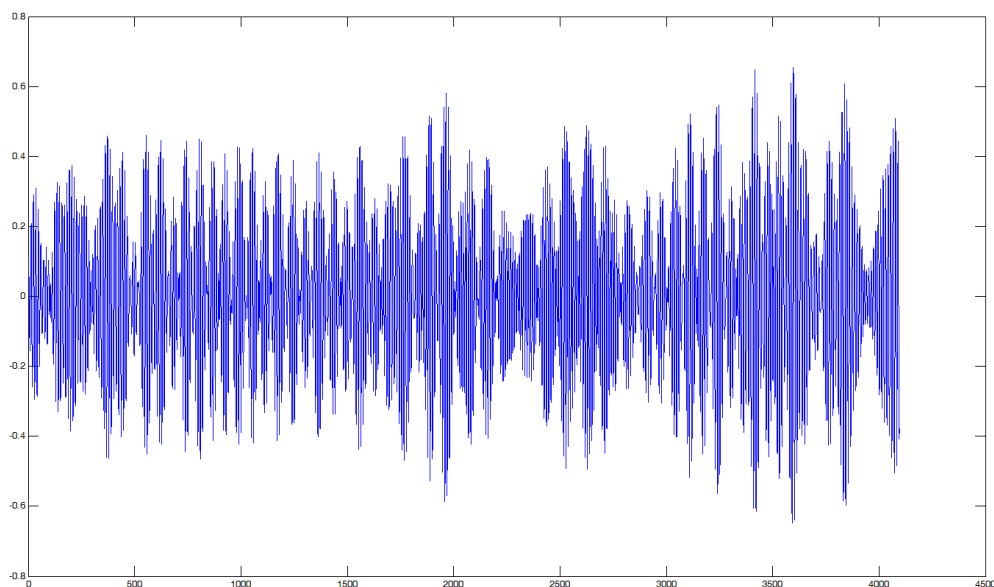
Per determinar mètodes eficients que permetessin detectar la presència de sons de xiulet en una senyal d'àudio era important conèixer les característiques de la senyal sonora de xiulet tant des del punt de vista temporal com freqüencial.

La caracterització dels xiulets era, per tant, necessària per tal d'enfocar el problema des d'una perspectiva realista. D'especial interès era conèixer-ne el seu comportament freqüencial [10]. Per aquesta raó es va realitzar una gravació dels xiulets Fox 40 i Naysir N1B. A partir d'aquestes gravacions se'n va calcular la DFT per veure el comportament en el domini freqüencial.

### 2.3.1. Fox 40

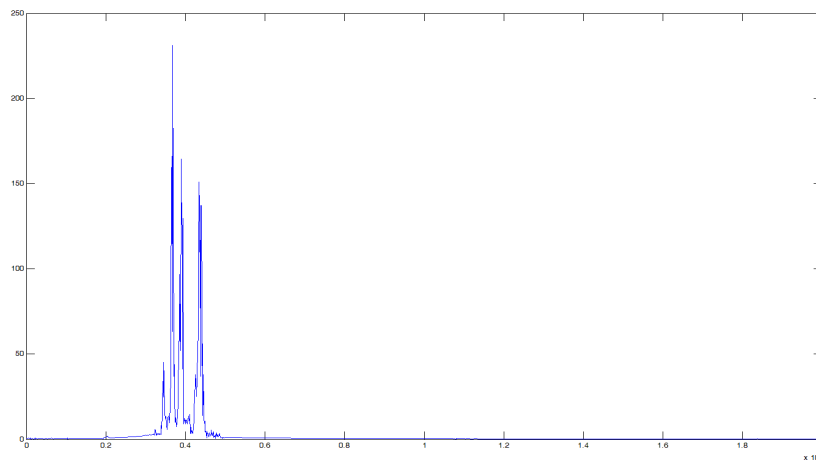
Tal i com ja s'ha comentat, el xiulet model Fox 40 és el model més emprat en l'arbitratge de futbol europeu. El seu preu ronda els 5 € i no mostra problemes davant situacions d'alta humitat degut a que no empra cap bola de cautxú en la cavitat interior.

En la Fig. 2.4 es mostra la forma d'ona d'un senyal d'àudio que conté un so de xiulet. En aquest cas concret es tracta d'un xiulet Fox 40 enregistrat sense so ambient, a una freqüència de mostrejat de 44.100 Hz i en mode monocal. El segment de so representant correspon a un fragment de 4096 mostres.



**Fig. 2.4** Fragment d'àudio de 4096 mostres d'un Fox 40.  
(Amplitud de les mostres en l'eix vertical, número de mostra en l'eix horitzontal)

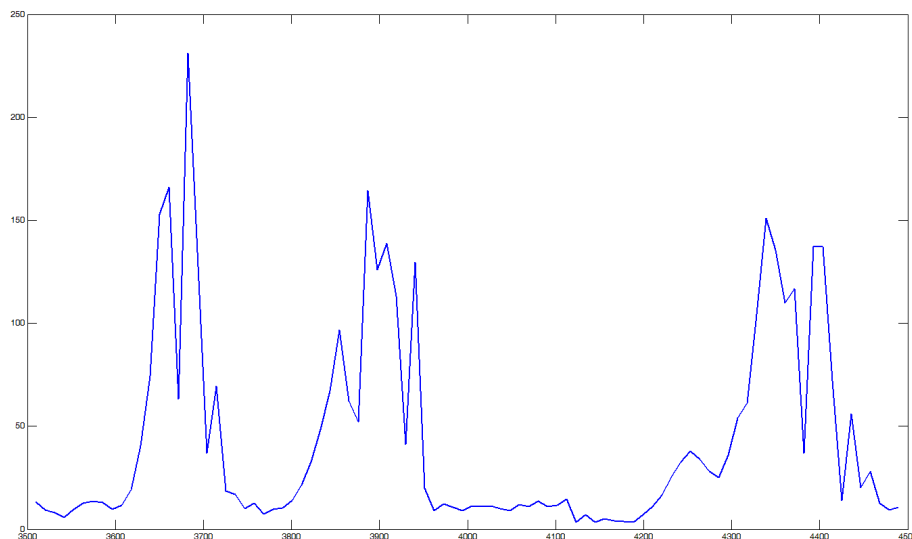
El mateix fragment de xiulet analitzat en el domini freqüencial revela la presència de components d'alta freqüència en regions espectrals situades entre els 3,5 kHz i els 4,5 kHz, tal i com mostra la Fig. 2.5.



**Fig. 2.5** DFT de 0 a 20 kHz d'un fragment de so d'un xiulet Fox 40 sense soroll ambient.

(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 10,8 Hz/mostra)

Més concretament podem observar-ho en la Fig. 2.6 en la qual es representa el mòdul de la DFT de 3,5 kHz a 4,5 kHz que s'ha realitzat sobre el mateix segment de so. La resolució espectral emprada [10] és de 10,8 Hz per mostra.



**Fig. 2.6** DFT de 3,5 kHz a 4,5 kHz d'un fragment de so d'un xiulet Fox 40 sense soroll ambient.

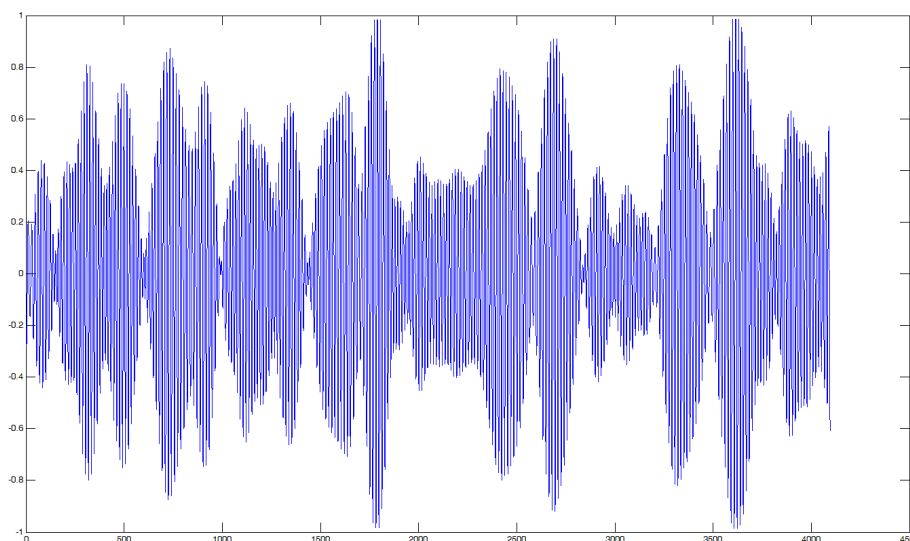
(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 10,8 Hz/mostra)

Es pot observar com el xiulet model Fox 40 emet un so que presenta tres pics de potència entorn les freqüències de 3650, 3900 i 4350 Hz. De totes maneres, tal i com s'explicarà més endavant, el xiulet model Fox 40 no produeix sempre un so amb exactament les mateixes freqüències sinó que aquestes poden variar, tot i que mantenint-se dintre el marge de 3,5 kHz a 4,5kHz.

### 2.3.2. Naysir N1B

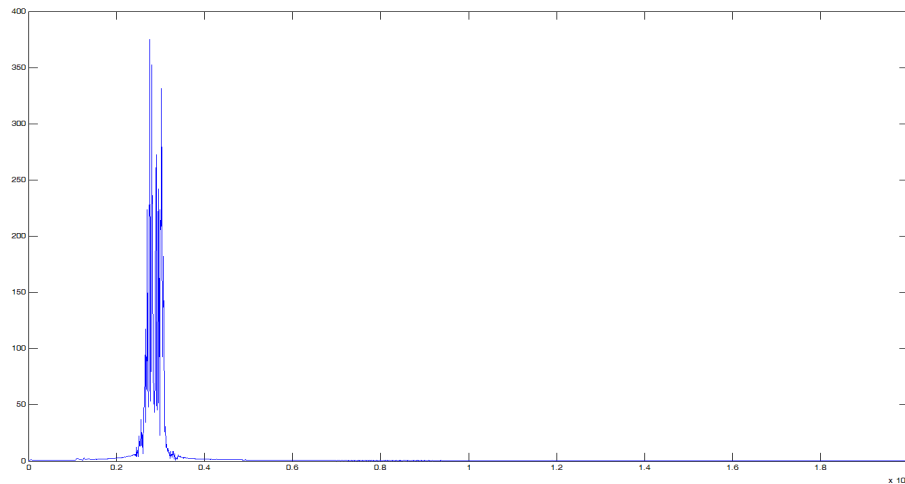
El xiulet Naysir N1B és un xiulet amb bola de cautxú a la cavitat interior i que té un preu inferior als 2 €. Pel fet d'emprar bola en la cavitat interior té tots els problemes propis que s'han comentat anteriorment. Malgrat tot, és un dels xiulets de bola més fiables i més estesos en ús, fet segurament degut a que produeix un so més agradable que el del Fox 40.

La Fig. 2.7 reproduïx un fragment de 4096 mostres d'un so de xiulet Naysir enregistrat en condicions absents de soroll ambient, a una freqüència de mostrejat de 44.100 Hz i en mode monocanal.



**Fig. 2.7** Fragment d'àudio de 4096 mostres d'un Naysir N1B.  
(Amplitud de les mostres en l'eix vertical, número de mostra en l'eix horitzontal)

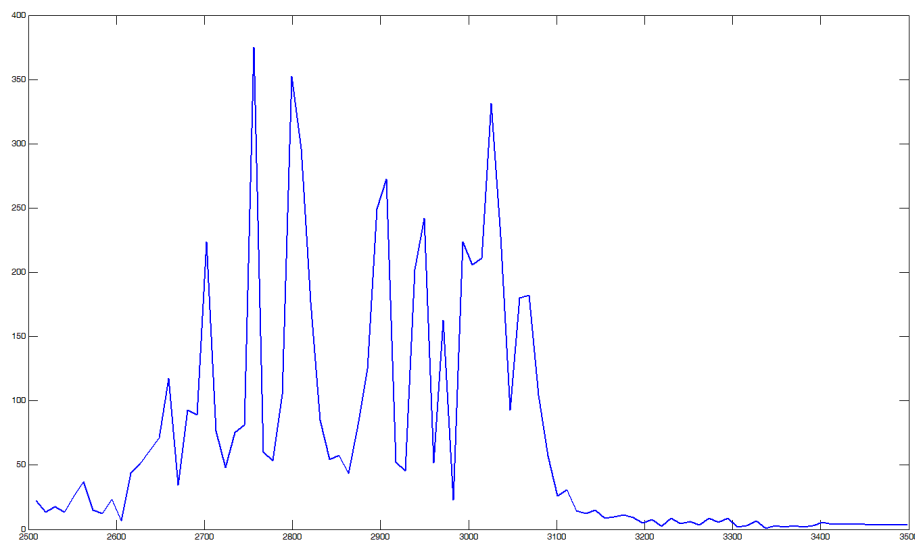
El mateix fragment de xiulet analitzat en el domini freqüencial revela la presència de components d'alta freqüència en regions espectrals situades entorn els 2,9 kHz, tal i com es pot observar en la Fig. 2.8.



**Fig. 2.8** DFT de 0 a 20 kHz d'un fragment de so d'un xiulet Naysir N1B sense soroll ambient.

(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 10,8 Hz/mostra)

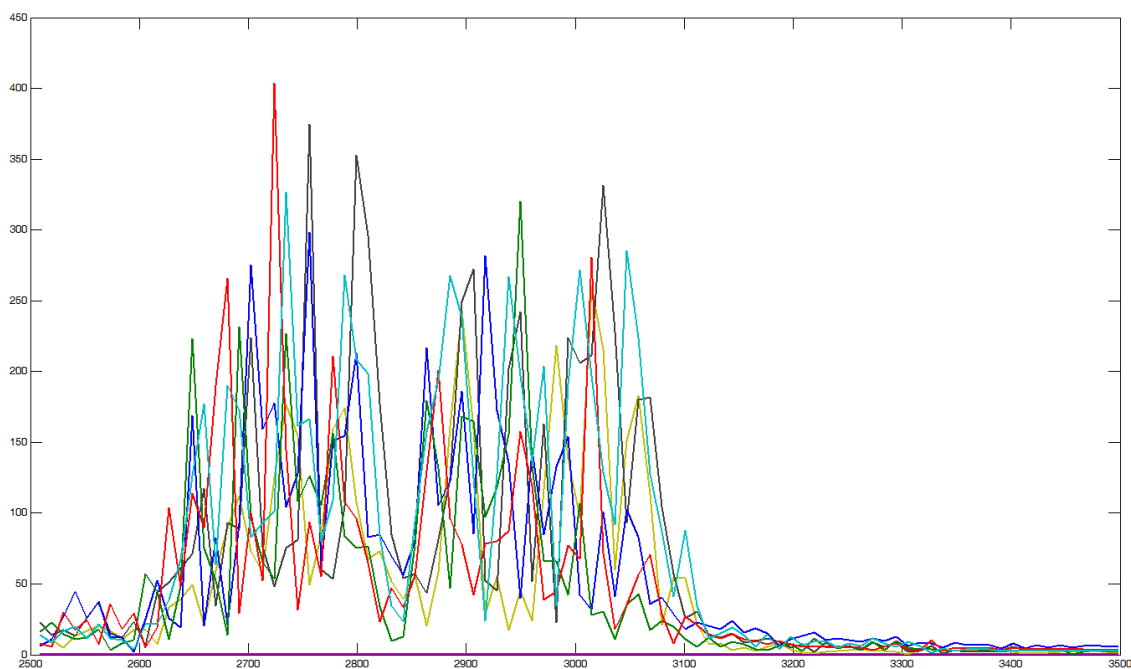
El comportament freqüencial de dit xiulet es pot observar en la Fig 2.9. Es representa el mòdul de la DFT de 2,5 kHz a 3,5 kHz que s'ha realitzat sobre el mateix segment de so de 4096 mostres. La resolució espectral emprada és de 10,8 Hz per mostra.



**Fig. 2.9** DFT de 2,5 kHz a 3,5 kHz d'un fragment de so d'un xiulet Naysir N1B sense soroll ambient.

(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 10,8 Hz/mostra)

Tal i com es pot observar, el xiulet model Naysir N1B no presenta una forma espectral definida sinó que crea un conjunt de pics de potència al voltant de la freqüència de 2900Hz. De fet, si es mostren superposades les DFTs de diferents segments del so d'un xiulet Naysir N1B s'evidencia que no hi ha una forma definida semblant entre elles.



**Fig. 2.10** DFTs superposades de diversos fragments de so d'un xiulet Naysir N1B sense soroll ambient.

(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 10,8 Hz/mostra )

Tal i com ha quedat constatat, les característiques tècniques dels dos models de xiulet analitzats tenen diferències notòries sobretot pel que fa al seu comportament freqüencial. Aquestes diferències es deuen sobretot a que un dels models és un xiulet de bola (Naysir N1B) i en canvi l'altre és sense bola en la cavitat interior (Fox 40). D'aquesta manera, per detectar cadascun dels xiulets, es va pensar en la possibilitat d'haver d'emprar una solució diferent per cadascun dels xiulets donada la diversitat de les seves característiques tècniques.

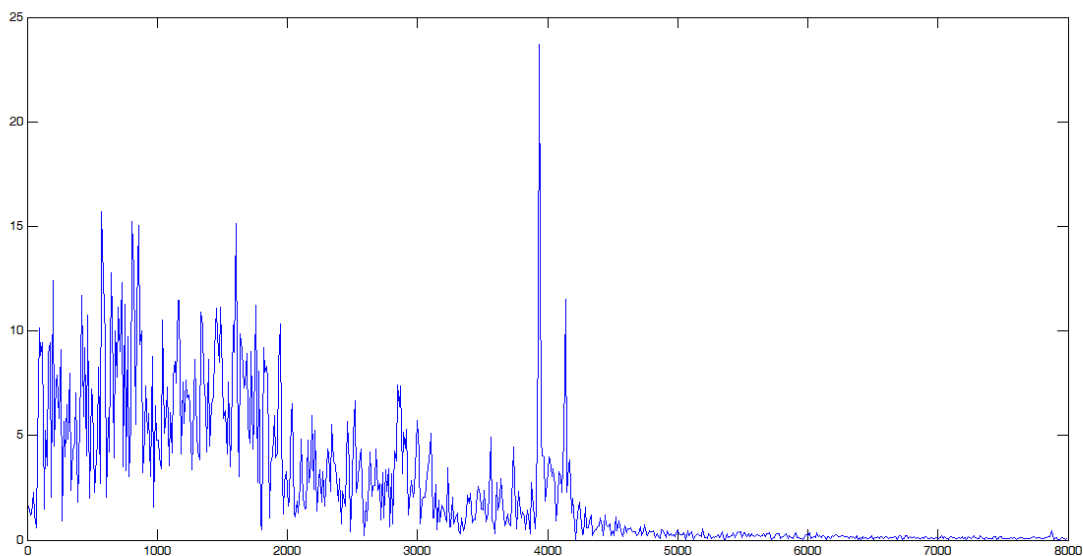
Donat el fet que la majoria d'àrbitres prefereix l'ús del xiulet Fox 40 per la seva fiabilitat i característiques tècniques i que aquest mateix xiulet és el xiulet oficial de la major part de competicions de futbol europeu (UEFA Europa League) i mundial (FIFA World Cup) es va decidir prioritzar la detecció del xiulet model Fox 40, deixant la detecció del model de xiulet Naysir N1B per una segona fase.

Tenint en compte les característiques tècniques del Fox 40, a continuació es presenten les primeres línies de treball emprades per tal d'intentar detectar la presència de xiulets de l'àrbitre ens vídeos esportius.

## 2.4. Primers mètodes emprats per la detecció de xiulets d'àrbitre

Tal i com s'ha observat en l'apartat anterior, el xiulet Fox 40 empra tres freqüències principals que defineixen el seu so dins la banda que va dels 3,5 kHz als 4,5 kHz. D'acord amb aquesta característica, la primera aproximació per intentar detectar un xiulet en una pista d'àudio va consistir en establir un banc de filtres en la zona d'interès de tal manera que si la potència de la senyal en els filtre on suposadament el Fox 40 hauria de tenir la major potència era efectivament major que la mitja es donaria com a positiva la detecció.

El problema d'aquesta tècnica consistia en què molt sovint el soroll ambient podia repercutir en un augment en la potència en la zona d'interès que faria saltar el detector donant un fals positiu. En la Fig 2.11 es mostra la representació de la DFT d'un segment d'àudio de 4096 mostres en el qual no hi ha un xiulet d'àrbitre però que presenta una potència elevada en la zona d'interès.

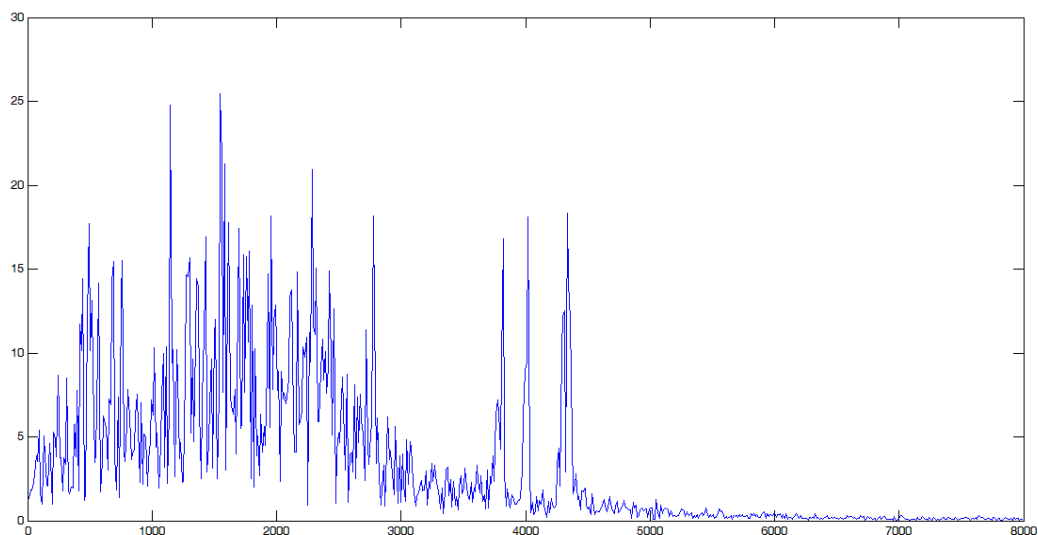


**Fig. 2.11** DFT d'un segment d'àudio sense so de xiulet amb presència de potència en la zona espectral d'interès.

(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

Tal i com es pot observar, hi ha un augment de la potència entorn els 4 kHz que faria pensar en un xiulet d'àrbitre quan en realitat es tracta de soroll ambient, donant-se per tant un fals positiu. Dit soroll ambient pot ésser causat

per trompetes d'animació, xiulets del públic, anuncis acústics que incorporen tons musicals, etc. A continuació, en la Fig 2.12, es representa la DFT d'un segment de 4096 mostres del mateix àudio que l'anterior en el que també hi ha un augment de potència entorn els 4 kHz i que sí que es tracta d'un xiulet d'àrbitre.



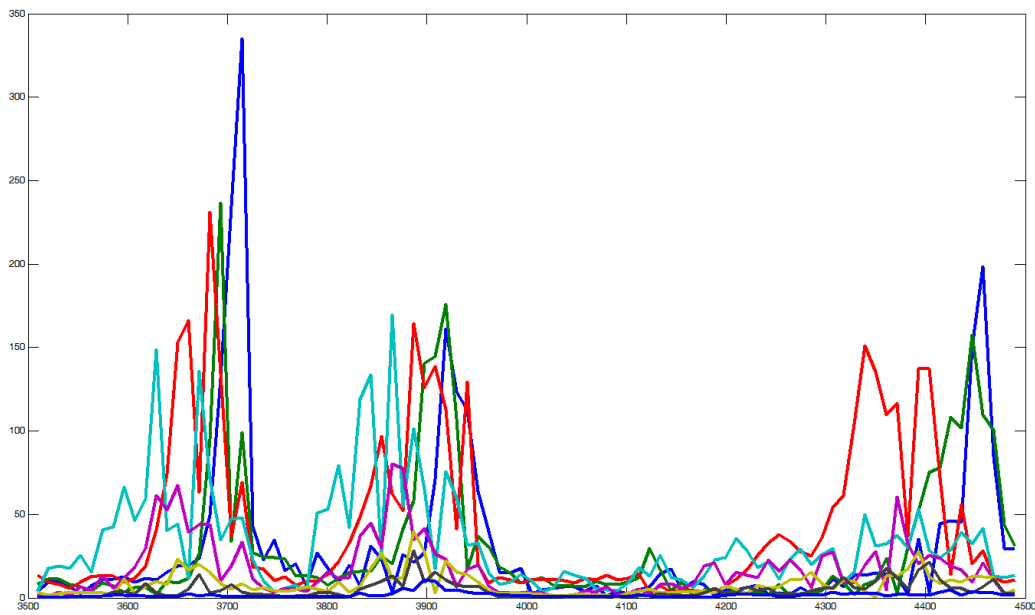
**Fig. 2.12** DFT d'un segment d'àudio amb presència de so de xiulet.  
(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

Tal i com es pot observar, la potència d'aquest segment entorn els 4 kHz és similar a la del segment anterior però així com aquest últim es tracta d'un xiulet, l'anterior reflectia simplement un so ambient.

Un segon mètode emprat va ser realitzar filtres de 50 Hz en les tres bandes d'interès en les quals el Fox 40 presenta cadascun dels tres pics de potència corresponents als que s'havien detectat en la caracterització del xiulet: 3650, 3900 i 4350 Hz.

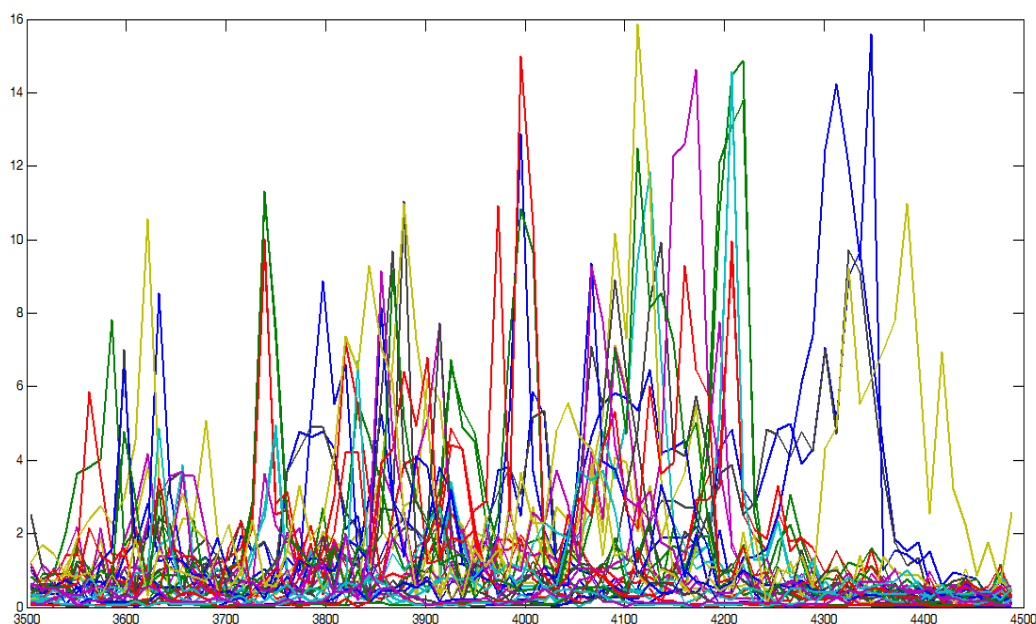
El problema d'aquesta tècnica va ser el fet que els tres pics de potència del Fox 40 podien variar de freqüència entre els diferents xiulets i fins i tot en un mateix xiulet. Per tal d'evidenciar aquest fet es mostra a continuació la superposició de les DFTs corresponents a diversos segments d'àudio d'un sol so gravat de xiulet.





**Fig. 2.13** DFTs superposades de diferents segments d'un so de xiulet Fox 40.  
(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

La Fig. 2.13 representa un sol so de xiulet en el qual s'han fet segments de 4096 mostres i se n'ha calculat cada una de les DFT. Es pot observar com hi ha un desplaçament en el domini freqüencial dels tres pics del so del xiulet. Tot i així encara es podria pensar que aplicar un filtre amb tres bandes de pas per detectar la presència de xiulet seria una solució factible. Però aquest desplaçament de freqüència dels tres pics de potència característics del Fox 40 es maximitza quan s'analitzen sons de xiulets de diferents partits. Per tal d'evidenciar el contrast la Fig. 2.14 mostra diferents segments superposats de diferents sons de xiulets de diversos partits de la LPF espanyola.



**Fig. 2.14** DFTs superposades de diferents xiulets Fox 40.

*(Eix horitzontal en Hz, potència en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)*

Tal i com es pot observar, la freqüència dels tres pics de potència del Fox 40 varien al llarg del temps en un mateix so de xiulet. Per aquesta raó, l'ús d'un banc de filtres no podia donar bon resultat.

Una tercera alternativa va ser buscar en la zona d'interès si es donava una concentració de potència en tres bandes freqüencials. La idea era que si la potència es concentrava entorn a tres freqüències (no calia que fos unes freqüències concretes sinó que es donessin tres concentracions diferents) llavors el sistema detectaria un xiulet i en cas contrari, no.

El problema d'aquesta tècnica va ser la irregularitat de situacions en què es donava el xiulet en els partits de futbol. És a dir, el soroll ambiental sovint feia que també es trobessin concentracions de potència en la zona d'interès de tal manera que el nombre de falsos positius era molt elevat.

Un cop comprovat que aquests tres mètodes no donaven resultats satisfactoris es va optar per aplicar el mètode d'anàlisi de components principals (Principal Component Analysis) conegut entre d'altres usos per les seves aplicacions en el reconeixement d'objectes. Si el mètode de PCA era capaç de detectar allò comú als objectes complexos (com una cara humana) en diverses situacions també hauria de ser capaç de detectar l'estructura fonamental d'un xiulet.

## **CAPÍTOL 3. ANÀLISI DE COMPONENTS PRINCIPALS (PCA)**

En aquest capítol es presenta el mètode emprat per a la detecció de xiulets de l'àrbitre en aquest treball. Vista la dificultat en establir uns paràmetres característics per mesurar la presència de xiulet arbitral, es va haver de pensar en un mètode que fos capaç de superar les dificultats trobades en un inici.

Després d'un estudi de diferents sons de xiulets de diversos partits de futbol, es va poder observar que eren reconeixibles fàcilment per l'ull humà quan aquests eren representats en el domini freqüencial. Efectivament, la forma espectral presentava una figura molt semblant consistent en 3 pics de potència que podien variar de freqüència i amplitud. Però malgrat la vista era capaç de detectar un xiulet en veure'l representat en el seu espectre, no s'havia trobat cap paràmetre o conjunt de paràmetres concrets capaços de detectar-ne la seva presència, fet degut, en gran part, a la seva variació freqüencial i potencial.

Va ser en aquest moment quan es va pensar en l'ús de l'Anàlisi de Components Principals com a possible solució donat que aquesta tècnica és emprada en reconeixement d'objectes complexos i, per tant, si és capaç de reconèixer objectes complexos hi ha gran probabilitat que també sigui capaç de reconèixer xiulets d'àrbitres.

A continuació es presenta una aproximament teòric al mètode d'Anàlisi de Components Principals i la descripció dels seus procediments.

### **3.1. Descripció del mètode d'Anàlisi de Components Principals**

El mètode d'Anàlisi de Components Principals [11] pertany al camp de reconeixement de patrons [12][13] el qual s'aplica a la detecció automàtica de regularitats en les dades mitjançant l'ús d'algoritmes computacionals. A partir d'aquestes regularitats trobades en un conjunt de dades es poden realitzar tasques com ara la classificació de les dades en diferents categories de manera automatitzada.

El mètode d'Anàlisi de Components Principals [11] es pot definir com la projecció ortogonal de les dades que es volen analitzar sobre un espai lineal de dimensió inferior, conegut com el subespai principal, tal que la variança de les dades projectades és màxima. Equivalentment es pot definir també com la projecció lineal que minimitza el cost mig de la projecció, definit com el quadrat de la distància entre els punts de les dades i les seves projeccions.

L'Anàlisi de Components Principals consisteix en un procediment matemàtic que transforma un nombre de variables correlacionades en un nombre més petit de variables no correlacionades anomenades components principals.

L'objectiu de l'anàlisi de components principals és identificar la base més significativa per tornar a expressar un conjunt de dades. L'esperança és que aquesta nova base filtrarà el soroll i revelarà l'estructura oculta. Així, la tècnica de PCA és emprada com a mètode d'exploració per descobrir tendències en les dades.

L'aplicació del mètode d'Anàlisi de Components principals per detectar el xiulet Fox 40 s'explica sota la pretensió de trobar un patró comú en el so dels xiulets. És a dir, trobar, en un conjunt de sons de xiulets enregistrats de diferents partits de futbol, l'estructura fonamental que permetés, davant un nou partit de futbol, detectar la presència de xiulets d'àrbitre donada la seva similitud amb dita estructura fonamental prèviament coneguda.

El problema que havia aparegut en els primers mètodes per detectar el xiulet de l'àrbitre és que malgrat aquest presentava tres pics de potència en freqüències entre els 3,5 i 4,5 kHz, les freqüències de dits pics podien variar i també la seva amplitud, de tal manera que es feia difícil trobar les característiques que permetessin detectar-lo i no confondre'l amb un so ambient que presentés característiques similars.

Calia doncs trobar aquella estructura comuna a tots els xiulets que els diferenciava de qualsevol altre tipus de so ambient amb característiques similars. És precisament això el que fa el mètode d'Anàlisi de Components Principals, trobar l'estructura comuna en un conjunt de dades.

Un cop realitzada l'extracció de dita estructura comuna a partir d'un conjunt de sons de xiulets de diferents partits de futbol es podria dir si una nova mostra de so d'un partit de futbol és o no un xiulet d'àrbitre segons si es pot reconstruir o no a partir de l'estructura comuna. La reconstrucció mai seria perfecte, però calculant-ne l'error i establint un llindar es podria decidir si el cas analitzat pertanyia a un xiulet o bé a un so ambient.

### 3.2. Procediments del PCA

L'anàlisi de components principals parteix d'un conjunt de dades organitzats en una matriu de M vectors de dimensions N cadascun d'ells. El PCA parteix de la base segons la qual qualsevol vector de M components es pot expressar en una base ortogonal de N vectors sense error.

$$X_N = \alpha_1 r_1 + \alpha_2 r_2 + \dots + \alpha_M r_M \quad (3.1)$$

El que pretén el PCA és trobar un conjunt de vectors ortogonals inferiors en nombre per tal d'expressar el vector  $X_M$  amb un error mínim.

$$X_K = \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_K v_K \quad (3.2)$$

El primer pas del PCA consisteix en el càlcul de la matriu de covariança de la matriu que conté el conjunt de dades.

$$R_{xx} = \frac{1}{M} \sum_{i=1}^M (X - \mu) \cdot (X - \mu)^T \quad (3.3)$$

La matriu de covariança serà una matriu quadrada de  $M \times M$ . A partir de la matriu de covariança es poden calcular els autovectors [11] i els autovalors [11]. Així, donada una matriu  $R$  quadrada, que en el nostre cas correspon a la matriu de covariança, es poden trobar els autovalors  $\lambda$  i autovectors  $v$  tals que es compleixi que:

$$R_{xx} \cdot v_k = \lambda_k \cdot v_k \quad \text{on} \quad k=1, \dots, M \quad (3.4)$$

D'aquesta manera haurem trobat els autovectors i autovalors de la matriu de covariança del conjunt de dades. El vector  $X_M$  pot ser expressat com a una combinació lineal dels seus autovectors normalitzats:

$$X_M = \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_N u_N \quad (3.5)$$

Els autovectors ens mostren les característiques més rellevants d'un conjunt de dades, ens diuen quines són les característiques més representatives respecte el conjunt de dades. D'aquesta manera, és possible ordenar els autovectors en una matriu segons el pes dels seus autovalors corresponents. Un cop ordenats els autovalors segons el valor dels seus autovalors corresponents, es normalitzen els autovectors per crear una base ortonormal:

$$u_k = \frac{v_k}{\sqrt{v_k^T v_k}} \quad (3.6)$$

La idea del PCA és emprar només els primers  $P$  coeficients (corresponents als autovectors amb autovalors majors), on  $P$  és menor que  $M$ :

$$\tilde{X} = \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_P u_P \quad \text{on} \quad \alpha_K = u_K^T \cdot X \quad (3.7)$$

Es pot demostrar que el valor mig de l'error al quadrat es pot calcular com:

$$E \left[ (X - \tilde{X})^T \cdot (X - \tilde{X}) \right] = \lambda_{P+1}^2 + \lambda_{P+2}^2 + \dots + \lambda_M^2 \quad (3.8)$$

## CAPÍTOL 4. PCA APLICAT A LA DETECCIÓ DE XIULETS

Si fins ara s'ha vist les característiques tècniques del xiulet que es vol detectar, la seva complexitat en quant a caracterització freqüencial, i els diversos mètodes emprats que no han assolit l'objectiu desitjat, a continuació es proposa una aplicació del mètode descrit en el capítol anterior. Es tracta d'aplicar el mètode d'Anàlisi de Components Principals en la detecció automàtica de xiulets d'àrbitre en partits de futbol de la LFP.

El sistema consta de dues parts diferenciades: una primera etapa d'entrenament i una segona d'anàlisi. La primera etapa, corresponent a l'entrenament del sistema, cerca l'estructura comuna o patró fonamental del so del xiulet que es vol detectar en un conjunt de sons de partits de futbol enregistrats prèviament en els quals hi ha presència d'un xiulet. La segona etapa, corresponent a l'anàlisi de la pista d'àudio, se centra en analitzar la pista d'àudio del partit en el qual es volen detectar els xiulets de l'àrbitre.

En aquest capítol es descriurà el funcionament de cada una de les parts de les dues etapes del sistema. Així, es veurà com l'etapa d'entrenament del sistema està formada per la compilació de DFTs [14] de xiulets mostra, el càlcul de la matriu de covariança i el càlcul d'autovectors i autovalors. L'etapa de l'anàlisi de la pista d'àudio consta de sis subetapes: entrada de la pista d'àudio per segments, càlcul de la DFT dels segments de 3,5 a 4,5 kHz, càlcul de les autocorrelacions del segment, reconstrucció del segment, criteri de decisió i el post-processat.

### 4.1. Diagrama de blocs

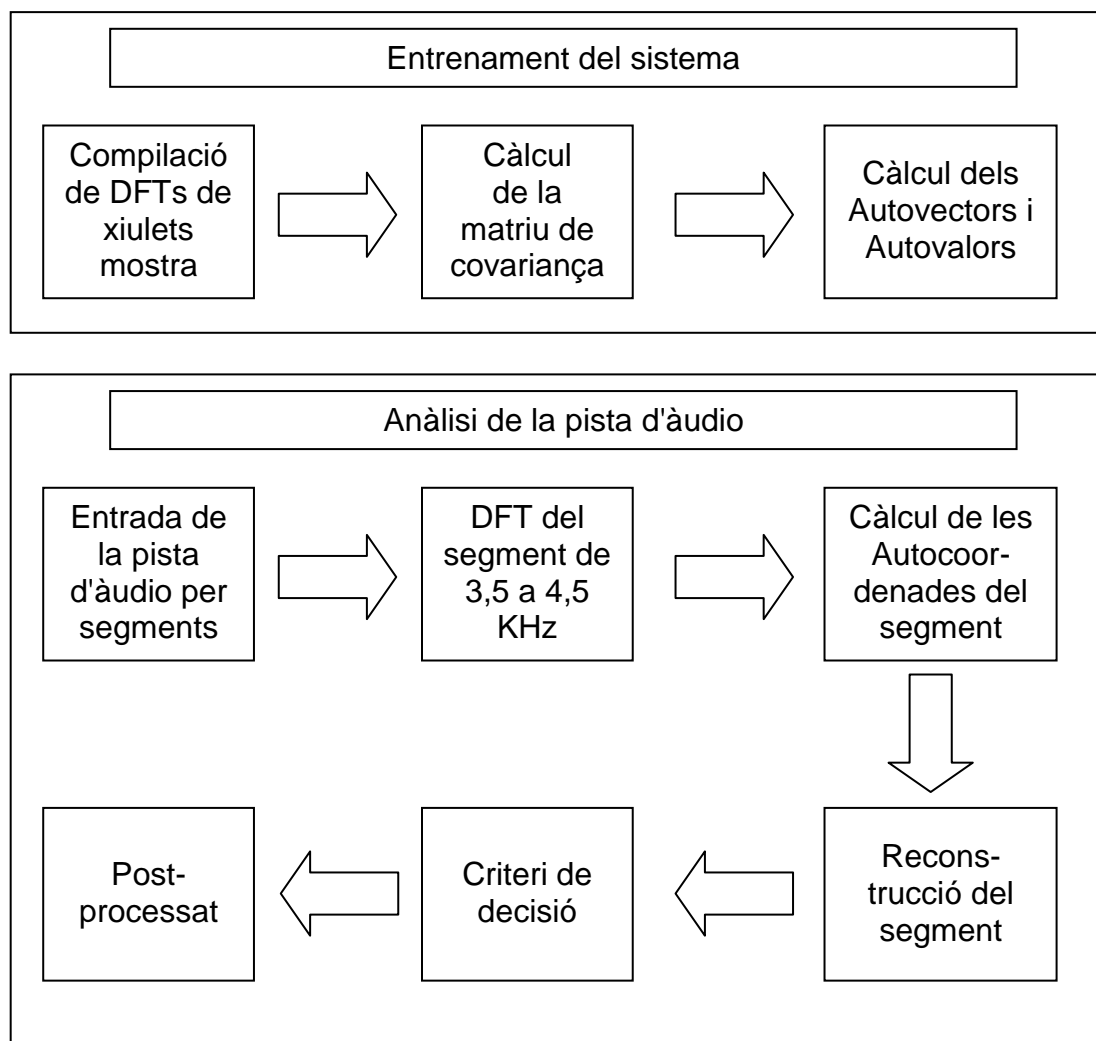
El diagrama de blocs presentat a continuació correspon a les etapes del sistema emprat per la detecció de xiulets en partits de la LFP espanyola. Representa les dues etapes que conformen el sistema, una etapa en la qual es compila la informació necessària del model de xiulet que es vol detectar i una segona etapa que cerca els sons del xiulet en la pista d'àudio que es vol analitzar.

La primera etapa té com a objectiu proporcionar un entrenament del sistema a partir de diferents sons de xiulets provinents de diferents partits de futbol. D'aquesta manera el sistema pot trobar el patró comú que defineix aquest grup de xiulets que al tractar-se del mateix model Fox 40 permetrà que en la segona etapa es pugui dur a terme una detecció de nous sons de xiulets de partits de futbol.

Aquesta primera etapa d'entrenament del sistema pot realitzar-se una sola vegada i a partir del patró comú trobat realitzar tots els anàlisis subsegüents de partits de futbol en els quals es vol detectar la presència dels xiulets de l'àrbitre.

La segona etapa és la que es realitza cada vegada que un partit de futbol ha de ser analitzat per detectar-ne els xiulets. En aquesta etapa s'analitza l'àudio per segments i es busca la seva distància amb el patró comú trobat en la primera etapa. Si aquesta distància (error de reconstrucció, definit al punt 4.1.8) és inferior a cert llindar el sistema conclourà que aquell segment que està analitzant és un xiulet i conclourà que no ho és en cas contrari.

La Fig. 4.1 es representa el diagrama de blocs del sistema amb les dues etapes diferenciades, l'entrenament del sistema i l'anàlisi de la pista d'àudio, cadascuna amb les seves subetapes concretades.



**Fig. 4.1** Diagrama de blocs de les dues etapes principals del sistema.

Un cop vist el conjunt del diagrama de blocs del sistema i l'especificitat de cada una de les dues grans etapes, en els següents apartats es descriurà cada una de les subetapes començant per les de l'etapa de l'entrenament dels sistema per prosseguir després amb les subetapes de l'etapa de l'anàlisi de la pista d'àudio.



#### 4.1.1. Compilació de DFTs de xiulets mostra

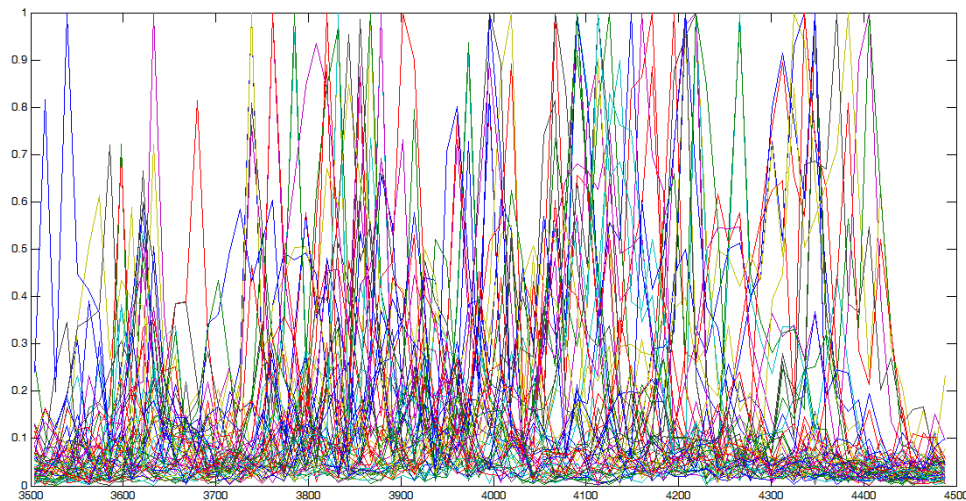
Per tal d'entrenar el sistema es precisen d'un cert nombre de xiulets reals de partits de la LFP que serviran de mostra. Aquests xiulets es tallen en segments de 4096 mostres. Aquest nombre de mostres es justifica a partir de la relació de compromís entre la necessitat que la DFT [10] tingui una definició suficient i la conveniència, en quant a estalvi de càrrega computacional, en què aquest nombre sigui el més petit possible.

La resolució espectral es calcula a partir de la freqüència de mostratge, que en les pistes d'àudio que es volen analitzar és de 48 kHz. Així:

$$\text{Resolució espectral} = \frac{\text{Freq. de mostratge}}{\text{Número de mostres del segment}} = 11,7 \text{ Hz/mostra} \quad (4.1)$$

De cadascun d'aquests xiulets mostra es calcularà la DFT per les freqüències que van de 3,5 a 4,5 KHz donat que és on es troben els tres pics de potència dels xiulets Fox 40. Per calcular només la DFT d'aquestes freqüències i estalviar així el cost computacional de calcular la DFT per tot l'ample de banda, el sistema utilitza la funció *goertzel* del programa Matlab. D'aquesta manera tindrem 85 mostres per xiulet amb una definició de 11,7 Hz per mostra.

La compilació de xiulets mostra emprada en el present projecte es compon d'una matriu de 52 DFTs de sons de xiulet, tal i com es mostra en la Fig. 4.2.



**Fig. 4.2** DFTs superposades dels segments dels diferents sons de xiulet de diferents partits de futbol emprats per l'entrenament del sistema.  
(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

El motiu pel qual s'agafa la DFT dels xiulets i no el domini temporal és perquè l'estructura oculta que s'espera que descobreixi el mètode d'Anàlisi de Components Principals es troba en el domini freqüencial. La idea de fons és que el sistema sigui capaç d'esbrinar quan en un segment d'àudio passat al domini freqüencial hi ha tres pics que no sempre seran iguals ni en amplitud ni en freqüència però que l'ull humà és capaç de distingir d'un altre so que pugui tenir una DFT semblant.

En el transcurs del projecte es va provar si el mètode PCA era capaç de detectar un segment de so de xiulet en el domini temporal però el resultat va ser negatiu: el nombre falsos positius era elevadíssim.

#### 4.1.2. Càlcul de la matriu de covariança

A partir de la matriu de 52 DFTs de sons de xiulet es calcula la matriu de covariança, la qual és una matriu quadrada de 52x52 mostres corresponents als valors de les covariances entre cadascuna de les 52 dimensions. Si la variança ens indica la variació respecte la mitja d'un conjunt de dades, la covariança mesura el valor esperat del producte de les desviacions respecte a la mitjana i indica el sentit de la correlació entre les variables.

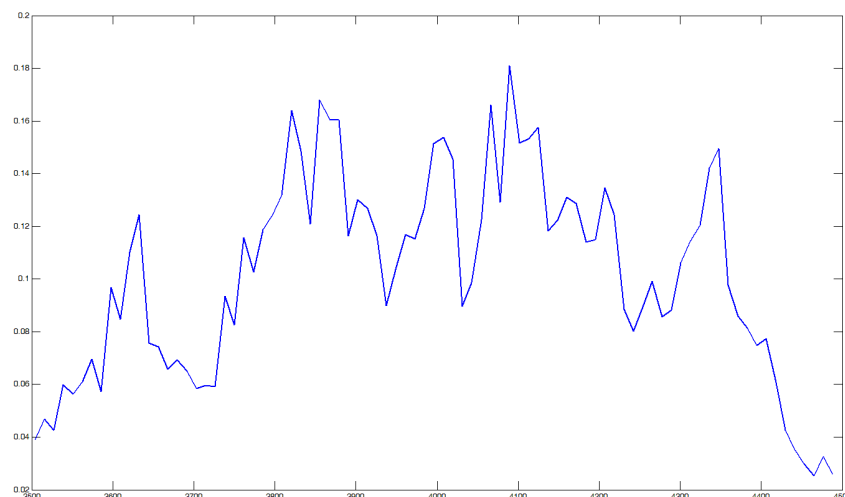
$$R_{xx} = \frac{1}{M} \sum_{i=1}^M (X - \mu) \cdot (X - \mu)^T = A \cdot A^T \quad (4.2)$$

On A és la matriu de dimensions 52x85 que conté la compilació de les DFTs de 3,5 kHz a 4,5 kHz dels sons de xiulets emprats per l'entrenament del sistema.

#### 4.1.3. Càlcul dels autovectors i autovalors

Un cop tenim la matriu de covariances es pot procedir a calcular els autovectors i autovalors d'aquesta matriu. Un cop calculats, els autovectors s'ordenen en una matriu segons el valor decreixent del seu respectiu autovalor que ens indica la importància que tenen respecte al conjunt d'autovectors. Així, els autovectors queden ordenats segons el valor del seu autovalor corresponent.

Una de les idees centrals del mètode d'anàlisi de components principals és reduir el nombre de dimensions inicials sense perdre la major part de la informació. Per a fer-ho cal que s'eliminin els autovectors que tenen un autovalor amb un valor inferior, per això els autovectors s'ordenen normalment segons el valor dels seus autovalors.



**Fig. 4.3** Autovector més representatiu (major valor associat d'autovalor)

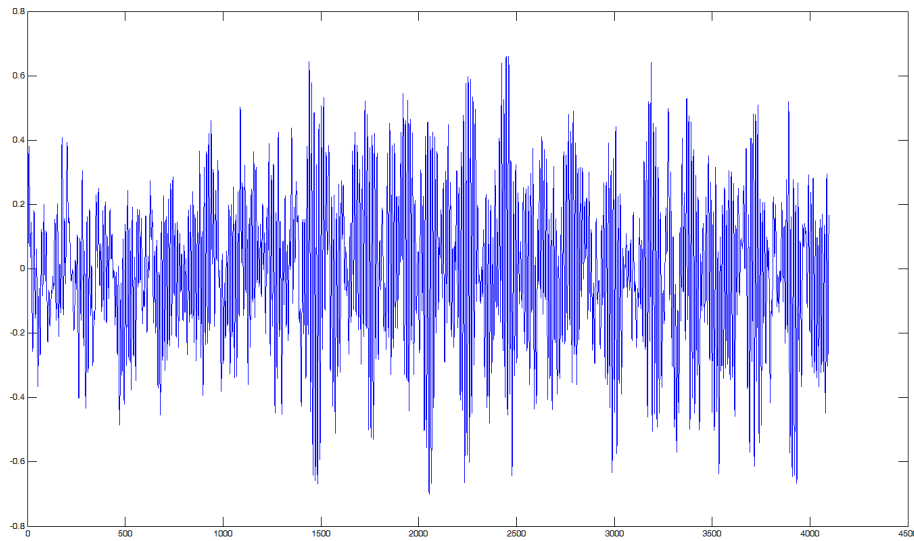
La Fig. 4.3 mostra l'autovector més representatiu (amb un valor associat d'autovalor més elevat) del nostre bloc d'entrenament del sistema. Aquest autovector és la base més comuna de totes les DFTs de sons de xiulets emprats en l'entrenament. A partir d'aquesta base es podria reconstruir qualsevol DFT de so de xiulet Fox 40 emprant una combinació lineal adequada de la resta d'autovectors amb un error de reconstrucció relativament petit.

Finalment els autovectors es normalitzen. Com que qualsevol autovector multiplicat per una constant donarà lloc a un altre autovector els autovectors s'acostumen a normalitzar per convenció.

#### **4.1.4. Entrada de la pista d'àudio**

Les pistes d'àudio corresponents a partits de la LFP espanyola acostumen a estar en format "wav" i a tenir una durada d'entre 2 i 3 hores ja que poden contenir l'escalfament dels jugadors i la mitja part també gravats. Això significa que el tamany de les pistes d'àudio dels partits acostuma a ser d'entre 1 i 2,5 GB.

Tenint present aquesta dada i el fet que es pretén detectar la presència o manca de xiulet en cada instant se segmenta la pista d'àudio en fragments el més petits possibles de manera que es minimitzi el cost computacional. Però la mida dels fragments també determina la resolució espectral que presentaran les DFTs que se'n calculen. Per aquesta raó s'estableix una relació de compromís i s'empren segments de 4096 mostres de manera que la resolució espectral, calculada a partir de la divisió entre la freqüència de mostreig de la pista d'àudio i el nombre de mostres, sigui de 11,72 Hz per mostra, donat que la freqüència de mostreig dels partits de futbol és de 48 kHz. La Fig. 4.4 representa un exemple de segment d'àudio de 4096 mostres.

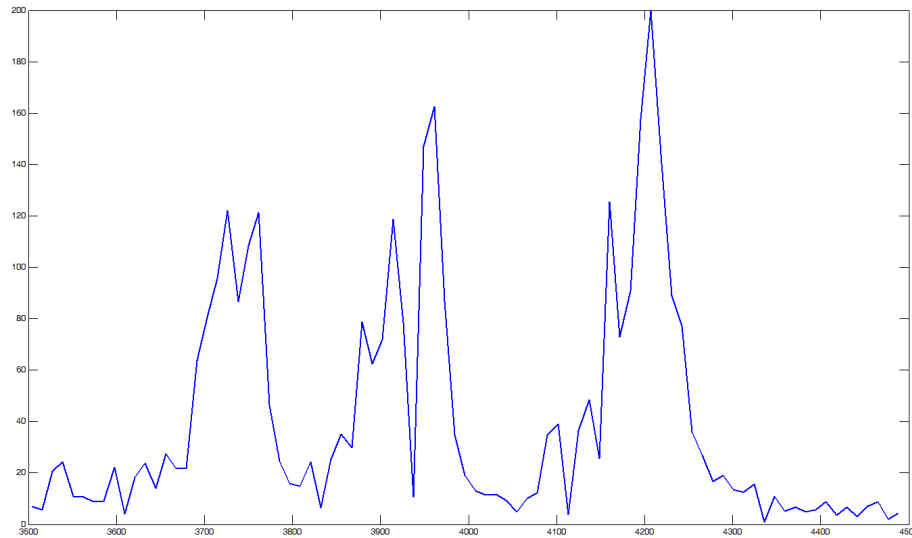


**Fig. 4.4** Fragment d'àudio de 4096 mostres d'un segment d'àudio amb presència de xiulet arbitral.

*(Amplitud de les mostres en l'eix vertical, número de mostra en l'eix horitzontal)*

#### 4.1.5. Càlcul de la DFT del segment

De la mateixa manera que s'ha fet amb els segments dels xiulets mostra de l'entrament del sistema, de cada segment que es vol analitzar de la pista d'àudio del partit se'n calcula la DFT de les freqüències que van de 3,5 KHz a 4,5 KHz emprant la funció *goertzel* del Matlab, evitant així el cost computacional de calcular-la per tot l'espectre freqüencial. El resultat són 85 mostres amb una definició d'11,7 Hz/mostra. La següent figura representa la DFT del segment de l'àudio representat en la figura anterior.



**Fig. 4.5** DFT del segment d'àudio mostrat en la Fig. 4.4.  
(Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

#### 4.1.6. Càlcul de les autocoordenades

El segment que s'està analitzant es multiplica per la matriu transposada d'autovectors aconseguida durant l'entrenament per obtenir-ne les corresponents autocoordenades. Les autocoordenades ens diran quina contribució de cada un dels autovectors aconseguits en l'entrenament del sistema aconseguix aproximar-se millor al segment original d'àudio que volem analitzar quan se'n calculi la reconstrucció a partir de les combinacions lineals dels autovectors.

El segment que s'està analitzant correspon a un segment d'àudio de 4096 mostres a les quals s'ha calculat la DFT de 3,5 kHz a 4,5 kHz. Un cop realitzada la DFT el segment resultant té 85 mostres. Per calcular les autocoordenades d'aquest segment es multiplica la matriu d'autovectors transposada, calculada en la subetapa anomenada càlcul d'autovectors i autovalors, pel segment de 85 mostres que es vol analitzar.

Sigui  $V$  la matriu d'autovectors calculada en l'etapa d'entrenament del sistema, amb un tamany de  $52 \times 85$ , i  $S$  el segment que es vol analitzar, les autocoordenades d'aquest segment,  $C$ , es poden calcular com segueix:

$$C_m = s_1 u_{m1} + s_2 u_{m2} + \dots + s_{(n-1)} u_{m(n-1)} \quad (4.3)$$

#### 4.1.7. Reconstrucció

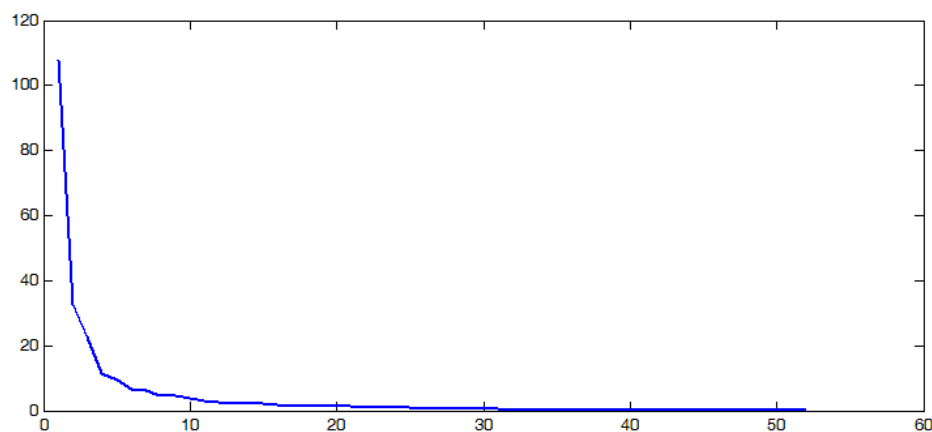
A partir de les autocoordenades corresponents al segment que s'està analitzant i els autovectors obtinguts en l'entrenament del sistema es realitza la reconstrucció del segment d'àudio que s'està analitzant.

La reconstrucció serà la suma de cada una de les contribucions dels autovectors que marquin les autocoordenades. Quant més s'assembli la reconstrucció al segment original més probabilitats tindrà de tractar-se d'un xiulet, ja que la reconstrucció haurà estat realitzada amb els autovectors compilats a partir dels xiulets mostra de l'entrenament del sistema.

El segment de reconstrucció resultant,  $R$ , tindrà un tamany de 85 mostres i es pot calcular com el sumatori de la multiplicació entre cadascuna de les 85 files de la matriu d'autovectors ( $V$ ) i els valors corresponents de les autocoordenades ( $C$ ):

$$R_n = u_{n1}c_1 + u_{n2}c_2 + \dots + u_{n(m-1)}c_{(m-1)} \quad (4.4)$$

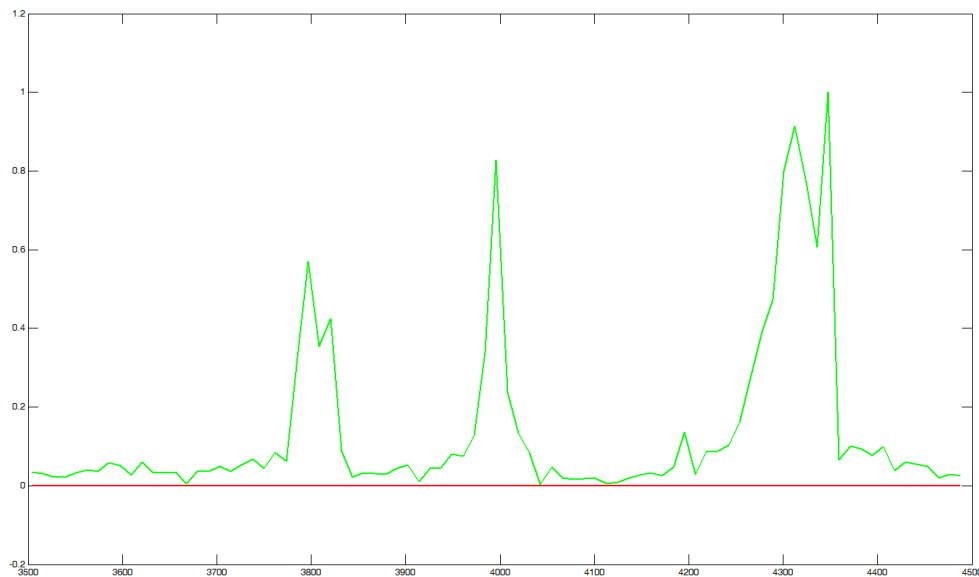
En el moment de la reconstrucció és quan és possible reduir el càlcul computacional reduint el nombre d'autovectors. Així, el nombre d'autovectors emprats ve determinat per  $M$  en la fórmula anterior. En el cas d'un sistema que ha emprat 52 sons de xiulet d'àrbitre per realitzar l'entrenament del sistema,  $M$  no podrà ser mai major de 52 però sí inferior. En qualsevol cas, degut a l'ordre en què s'han col·locat els autovectors (segons el seu autovalor), s'emprarien sempre els autovectors més representatius. La Fig 4.6 mostra la representació dels autovalors corresponents als autovectors.



**Fig. 4.6** Representació dels autovalors de la matriu d'entrenament del sistema

Si, per exemple, es vol reconstruir un xiulet que ja era present en l'entrenament del sistema, emprant tots els autovectors s'hauria d'esperar que l'error de reconstrucció fos despreciable. En canvi, si es redueix el nombre d'autovectors usats, deixant sempre els més significatius, l'error serà progressivament més gran.

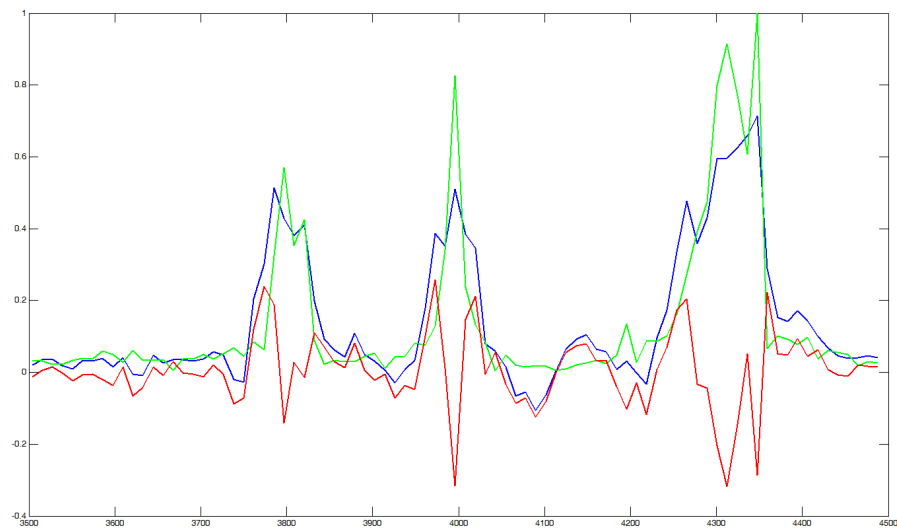
A continuació es mostra, en la Fig. 4.7, la reconstrucció d'un xiulet ja existent en el sistema d'entrenament primer emprant tots els autovectors i després reduint-ne el nombre usat a 5.



**Fig. 4.7** Reconstrucció de la DFT representada en la Fig 4.5 emprant tots els autovectors aconseguits en el sistema d'entrenament.

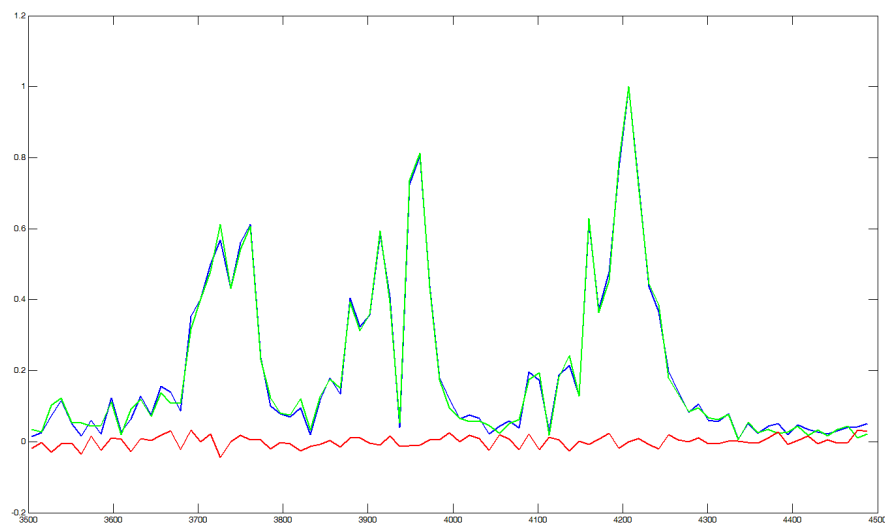
*(Reconstrucció representada en color verd, original en color blau, l'error de reconstrucció es representa en color vermell. Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)*

La reconstrucció (línea de color verda) tapa la línia blava que marcava la representació original de la DFT. La línia vermella indica la diferència entre les dues i per tant l'error, que en aquest cas és  $3 \cdot 10^{-13}$  (la manera de calcular l'error es troba descrita en el subapartat següent). Si reduïm el nombre d'autovectors emprats a 5 veurem com l'error augmenta a 5,95. Dita reconstrucció es troba representada en la Fig. 4.8.



**Fig. 4.8** Reconstrucció de la DFT representada en la Fig 4.5 emprant els 5 autovectors més representatius aconseguits en el sistema d'entrenament. (Reconstrucció representada en color verd, original en color blau, l'error de reconstrucció es representa en color vermell. Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

Mostrem ara la reconstrucció d'un xiulet que no era present en l'entrenament del sistema primer emprant tots els autovectors i després només els 22 més representatius i finalment una reconstrucció amb només 5 autovectors. La reconstrucció emprant tots els autovectors es mostra en la Fig 4.9.

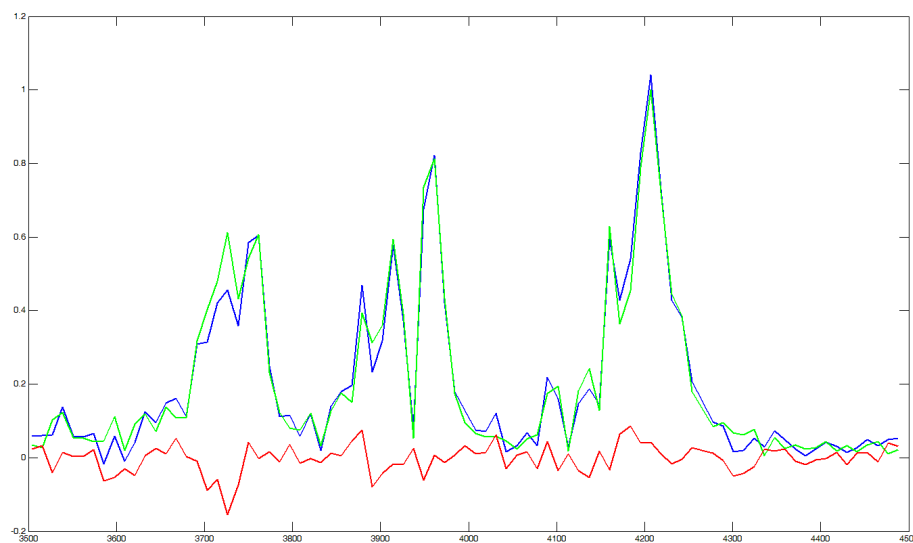


**Fig. 4.9** Reconstrucció d'un xiulet no present en el sistema d'entrenament emprant tots els autovectors. (Reconstrucció representada en color verd, original en color blau, l'error de reconstrucció es representa en color vermell. Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)



La línia de color blau representa la reconstrucció aconseguida a partir de la contribució de tots els autovectors segons marcaven les autocoordenades. La línia de color verd representa la mostra original del xiulet entrant. La línia vermella és la diferència entre cada un dels valors de la mostra original del xiulet entrant i la seva reconstrucció, és per tant l'error final de reconstrucció. En aquest cas, l'error final de reconstrucció (definit al punt 4.1.8) era de 1,10.

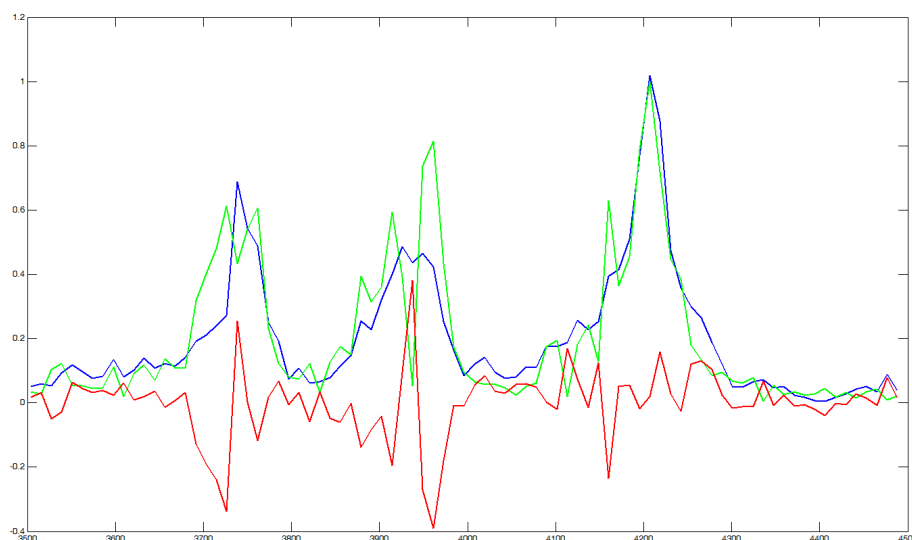
Si aquest mateix so de xiulet entrant l'intentéssim reconstruir només a partir dels 22 autovectors més representatius (d'un total de 52) la reconstrucció seria pitjor però encara força fidel a la original tal i com es mostra en la Fig. 4.10.



**Fig. 4.10** Reconstrucció de la DFT d'un xiulet no present en el sistema d'entrenament emprant els 22 autovectors més representatius.

*(Reconstrucció representada en color verd, original en color blau, l'error de reconstrucció es representa en color vermell. Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)*

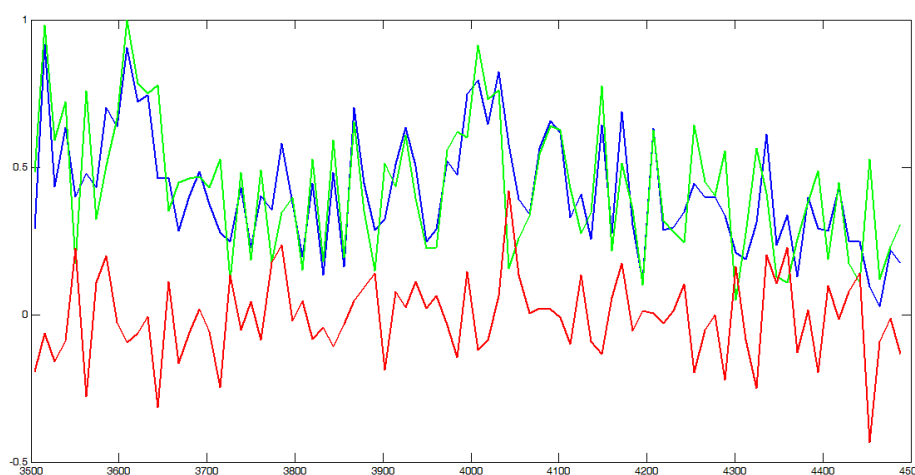
En aquest cas l'error final de reconstrucció hauria augmentat a 2,49. Si reduïssim a 5 el nombre d'autovectors emprats en la reconstrucció llavors l'error encara seria més gran. Així es mostra en la figura següent, l'error hauria augmentat a 6,02. La diferència es pot apreciar en la Fig 4.11.



**Fig. 4.11** Reconstrucció de la DFT d'un xiulet no present en el sistema d'entrenament emprant els 5 autovectors més representatius.

(Reconstrucció representada en color verd, original en color blau, l'error de reconstrucció es representa en color vermell. Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

A continuació es presenta el mateix procediment aplicat a un segment d'àudio de 4096 en el qual no hi ha xiulet d'àrbitre per veure quins resultats s'obtenen. En primer lloc es mostra el resultat de la reconstrucció emprant tots els autovectors. L'error que s'obté és de 9,27 i es pot veure en la Fig. 4.12.

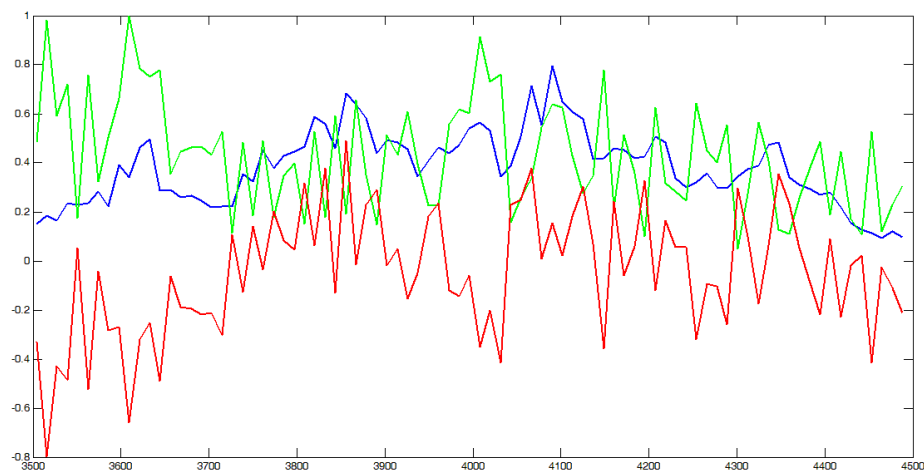


**Fig. 4.12** Reconstrucció de la DFT d'un segment d'àudio sense presència de xiulet emprant tots els autovectors.

(Reconstrucció representada en color verd, original en color blau, l'error de reconstrucció es representa en color vermell. Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

L'error en aquest cas és major fins i tot que l'error de reconstruir amb només 5 autovectors el segment que s'ha analitzat anteriorment i que sí contenia un xiulet d'àrbitre.

Si es realitza la reconstrucció amb només 5 autovectors l'error augmenta a 17,28. La representació d'aquesta reconstrucció i l'error respecte l'original es pot observar en la Fig. 4.13.



**Fig. 4.13** Reconstrucció de la DFT d'un segment d'àudio sense presència de xiulet emprant els 5 autovectors més representatius.

(Reconstrucció representada en color verd, original en color blau, l'error de reconstrucció es representa en color vermell. Eix horitzontal en Hz, potència en escala lineal en l'eix vertical. Resolució espectral de 11,7 Hz/mostra)

#### 4.1.8. Criteri de decisió

Un cop realitzada la reconstrucció, es calcula l'error com a la suma de les diferències (mostra a mostra) entre el segment de so que s'estava analitzant i la reconstrucció. S'entén que quan la reconstrucció hagi estat molt semblant al segment de so original aquest haurà de tractar-se d'un xiulet ja que tindrà una estructura similar.

Sigui  $E$  l'error que es vol calcular,  $R$  el segment de 85 mostres de la reconstrucció i  $S$  el segment que s'està analitzant, l'error pot calcular-se com:

$$E = \sum_{i=0}^{n-1} |R_i - S_i| \quad (4.5)$$

Diferents maneres de calcular l'error final van ser considerades i provades en el transcurs del projecte. En primer lloc es va provar de calcular l'error com a la suma de les diferències en valor absolut entre la reconstrucció i el segment original. Més tard es va provar de calcular l'error final com la suma de les diferències al quadrat igualment entre la reconstrucció i el segment original. També es va provar de calcular l'error final com en els dos casos anteriors però dividint-lo per la potència mitja del segment. Però aquests mètodes no ajudaven a distingir millor els segments que contenen un so de xiulet dels que no, ja que no augmentaven l'error de la reconstrucció dels segments sense xiulet respecte als segments amb xiulet.

#### **4.1.9. Post-processat**

El post-processat consisteix en dues etapes diferenciades. Una primera etapa comprova l'existència de tres pics de potència separats per una distància mínima de 100 Hz i amb una potència mínima de 0,2 vegades la màxima potència del segment estudiat. Els segments que havien estat detectats com a possibles sons de xiulet són ara descartats si no presenten tres pics diferenciats per les característiques que s'acaben de mencionar.

Els primers resultats van mostrar que el sistema era capaç de reconstruir DFTs de sons sense xiulet que tenien un o dos pics de potència amb un error de reconstrucció semblant a molts fragments que sí contenen un xiulet. Per aquesta raó es va pensar en una etapa de post-processat que eliminés aquests falsos positius. Aquesta etapa de post-processat no es podria haver aplicat sense la discriminació que ja fa el PCA doncs aquest discrimina una gran quantitat de fragments que presenten més de 2 pics de potència perquè no els pot reconstruir a partir dels autovectors.

Una segona etapa de post-processat consisteix en descartar els segments analitzats que han donat positiu fins ara però que per la seva proximitat temporal no poden representar dos esdeveniments d'interès diferenciats.

Donat que la freqüència de mostreig dels arxius d'àudio que contenen els partits de futbol és de 48 KHz, els segments que s'agafen per ser analitzats representen 0,085 segons d'àudio (4096 mostres / 48KHz). Per aquesta raó no es donen com a vàlids tots els segments analitzats que han donat positius després del criteri de decisió i de la primera etapa de post-processat ja que es pot donar el cas que se'n detecti algun que forma encara part del mateix xiulet.

Per altra banda, en un resum d'esdeveniments d'interès d'un partit de futbol, les jugades tenen una durada mínima i per tant, encara que en una jugada es detectin dos xiulets només caldrà marcar-ne el primer.

D'aquesta manera s'entén com a positiu en el criteri de decisió final només aquell segment analitzat que apareix com a positiu si l'últim positiu havia ocorregut anteriorment a una distància temporal de 10 segons o més.

## CAPÍTOL 5. PRESENTACIÓ DE RESULTATS

Un cop explicades cada una de les etapes del sistema i les seves parts es presenten a continuació els resultats obtinguts. En primer lloc es defineixen els conceptes matemàtics emprats per mesurar els resultats, com són la precisió, l'exhaustivitat i l'exactitud.

Un cop definits aquests termes es mostren els resultats en dos escenaris diferents. En el primer escenari es prepara el sistema per tal d'aconseguir una relació de compromís entre l'exhaustivitat, la precisió i l'exactitud. En el segon escenari es mostren els resultats quan el sistema tendeix a buscar la màxima exactitud, amb la reducció de la precisió que això suposa.

### 5.1. Precisió, exactitud i exhaustivitat

Per tal de presentar resultats del mètode de PCA aplicat a la detecció de xiulets per detectar esdeveniments d'interès en vídeos esportius s'empraran els termes de precisió (precision en anglès), exactitud (accuracy) i exhaustivitat (recall). Aquests termes empenen, al seu torn, els conceptes referents als falsos i veritables positius i negatius.

- Fals positiu (FP): quan el sistema considera que un segment analitzat conté un so de xiulet però realment no és així.
- Fals negatiu (FN): quan el sistema considera que un segment analitzat no conté un so de xiulet però sí que el conté en realitat.
- Veritable positiu (TP): quan el sistema considera que un segment analitzat conté un so de xiulet i realment és així.
- Veritable negatiu (TN): quan el sistema considera que un segment analitzat no conté un so de xiulet i realment és així.
- Precision [15] (precisió): es pot definir com la fracció d'elements de contingut que el mòdul de classificació ha identificat com a positius i que realment ho són. La precisió està inversament relacionada amb els falsos positius, i es pot calcular dividint els veritables positius entre la suma de veritables i falsos positius.

$$P = \frac{TP}{TP + FP}$$

(5.1)

- Accuracy [16] (exactitud): és definit com el grau de proximitat de les estimacions als valors reals. És el resultat de dividir la suma de

veritables positius i negatius entre la suma de veritables i falsos positius i negatius.

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.2)$$

- Recall [17] (exhaustivitat): es pot definir com la fracció dels elements de contingut que són realment rellevants a una categoria, que són reconeguts com a tals pel mòdul de classificació (la resta són falsos negatius). L'exhaustivitat està inversament relacionada amb els falsos negatius i es pot calcular com la divisió entre els veritables positius i la suma de veritables positius i falsos negatius

$$R = \frac{TP}{TP + FN} \quad (5.3)$$

Per manejar el trade-off entre el l'exhaustivitat i la precisió s'acostuma a emprar un llindar. Un alt llindar produeix una alta precisió però en detriment de l'exhaustivitat, en canvi, un baix llindar provoca una disminució de la precisió però una millora en l'exhaustivitat.

És per això que a continuació es presenten dos casos en els quals els resultats varien segons si es vol primar l'exhaustivitat o la precisió. Si es desitja reduir al màxim el nombre de falsos negatius tendint a una exhaustivitat del 100%, llavors cal que l'error de reconstrucció que s'admet sigui ampliat, cosa que produeix que el nombre de falsos positius augmenti de manera que disminueix la precisió.

En el cas de la detecció automàtica d'esdeveniments d'interès en vídeos esportius és discutible en quina mesura és necessari que cap esdeveniment d'interès quedi sense etiquetar. La pregunta és fins a quin punt és indispensable que no hi hagi cap xiulet que es quedi sense ésser detectat, és a dir, l'absència de falsos negatius. Sobretot si es té en compte que la detecció d'esdeveniments d'interès ve acompanyada, tal i com s'ha explicat anteriorment, de l'anàlisi d'altres descriptors d'àudio i anàlisi de vídeo per detectar els esdeveniments d'interès.

## 5.2. Resultats

Tal i com s'ha mostrat cal arribar a una relació de compromís entre l'exhaustivitat i la precisió. Amb aquest objectiu es presenten els resultats en escenaris diferents: un amb el llindar d'error més baix que porta a una millor precisió i un segon escenari amb un llindar major que porta a una exhaustivitat molt alta.

Es tracta de la detecció duta a terme sota aquests dos criteris diferents sobre un mateix partit de futbol de primera divisió. La mostra parcial del resultat del primer dels casos es representa en la Taula 5.1., on resten detallats tots els segments que han resultat positius després de l'anàlisi. Els que corresponen a un so de xiulet són marcats com a veritables positius mentre que els que no corresponen a cap so de xiulet són marcats com a falsos positius.

Els resultats s'han realitzat emprant, en l'etapa d'entrenament del sistema, 52 sons de xiulet de diferents partits de futbol de la LPF espanyola dels quals s'ha extret una matriu d'autovectors de tamany 52x85.

**Taula 5.1.** Mostra de resultats de l'anàlisi d'un partit de LPF espanyola.

Mostra inicial	Mostra final	Segment	Error	Temps	Veritables positius	Falsos positius
1187841	1191936	291	3,9	0:00:25		1
1859585	1863680	455	3,7	0:00:39		1
5787649	5791744	1414	3,6	0:02:01		1
7434241	7438336	1816	3,8	0:02:35		1
9060353	9064448	2213	3,6	0:03:09		1
10895361	10899456	2661	3,6	0:03:47		1
12574721	12578816	3071	3,4	0:04:22	1	
13279233	13283328	3243	3,7	0:04:37		1
15839233	15843328	3868	3,1	0:05:30	1	
16482305	16486400	4025	3,4	0:05:43	1	
18411521	18415616	4496	4,0	0:06:24	1	
19517441	19521536	4766	4,0	0:06:47		1
21958657	21962752	5362	3,9	0:07:37		1
25174017	25178112	6147	3,6	0:08:44		1
26116097	26120192	6377	3,3	0:09:04	1	
29646849	29650944	7239	3,4	0:10:18		1
31801345	31805440	7765	3,7	0:11:03		1
32481281	32485376	7931	4,0	0:11:17		1
34529281	34533376	8431	3,7	0:11:59		1
35143681	35147776	8581	3,4	0:12:12		1
36749313	36753408	8973	4,0	0:12:46	1	
37957633	37961728	9268	3,8	0:13:11		1
40148993	40153088	9803	3,6	0:13:56		1

44113921	44118016	10771	3,3	0:15:19		1
44937217	44941312	10972	3,7	0:15:36	1	
45633537	45637632	11142	3,6	0:15:51	1	
46833665	46837760	11435	3,5	0:16:16		1
47878145	47882240	11690	3,8	0:16:37		1
49528833	49532928	12093	3,1	0:17:12		1
52428801	52432896	12801	2,7	0:18:12		1
56717313	56721408	13848	3,6	0:19:42		1
57856001	57860096	14126	3,9	0:20:05		1
59256833	59260928	14468	3,9	0:20:35	1	
60964865	60968960	14885	3,2	0:21:10		1
62550017	62554112	15272	3,5	0:21:43		1
67411969	67416064	16459	3,9	0:23:24		1
67907585	67911680	16580	3,7	0:23:35	1	
71700481	71704576	17506	3,6	0:24:54		1
72945665	72949760	17810	3,0	0:25:20	1	
75513857	75517952	18437	3,7	0:26:13		1
76935169	76939264	18784	3,0	0:26:43		1
79622145	79626240	19440	3,7	0:27:39	1	
80281601	80285696	19601	3,8	0:27:53		1
81465345	81469440	19890	3,4	0:28:17		1
82292737	82296832	20092	3,8	0:28:34		1
84062209	84066304	20524	3,8	0:29:11		1
84594689	84598784	20654	3,2	0:29:22		1
85151745	85155840	20790	2,8	0:29:34	1	
86093825	86097920	21020	3,9	0:29:54		1
88698881	88702976	21656	3,3	0:30:48		1
92368897	92372992	22552	3,8	0:32:04	1	
94494721	94498816	23071	3,5	0:32:49	1	

Els resultats en termes de classificació són els següents:

- Accuracy: 99,8%
- Precision: 28,8%
- Recall: 88,2%

Tal i com es pot observar, en aquest primer cas l'exhaustivitat (recall) no arriba al 100% la qual cosa significa que s'han deixat de detectar alguns sons de xiulets.

Per esmenar aquesta situació i aconseguir una exhaustivitat el més elevada possible cal admetre un llindar. Els resultats que es presenten a continuació corresponen a l'augment del llindar de l'error de reconstrucció admès per tal



d'aconseguir una exhaustivitat del 100% en l'anàlisi del mateix partit de futbol emprat pels resultats anteriors:

- Accuracy: 99,5%
- Precision: 12,6%
- Recall: 100%

Tal i com es pot observar, el cost d'aconseguir un 100% d'exhaustivitat és una disminució considerable de la precisió. Cal notar també la importància d'haver assolit el 100% d'exhaustivitat encara que suposi un resultat menys precís, ja que un 100% d'exhaustivitat indica que el mètode és capaç de detectar tots els xiulets de l'àrbitre.

Per altra banda, els resultats mostrats fins al moment no manifesten completament el comportament del mètode d'Anàlisi de Components Principals pel que fa a detecció de xiulets de l'àrbitre. És a dir, els resultats són computats tenint en compte els xiulets però el PCA tal i com s'ha aplicat en el sistema analitza segments d'àudio, de tal manera que un sol so de xiulet pot ser detectat més d'una vegada (recordem que cada segment correspon a 0.085 s).

Si tenim en compte quan el PCA és capaç de detectar un xiulet en un segment determinat d'àudio llavors la precisió en l'anàlisi del mateix partit que s'ha emprat en els resultats anteriors arriba al 49,8%.

## CAPÍTOL 6. LÍNIES FUTURES DE TREBALL

En el transcurs del desenvolupament del sistema per detectar els xiulets de l'àrbitre en vídeos esportius en el context de la extracció automàtica d'esdeveniments d'interès, s'han anat veient les dificultats i potencialitats de l'aplicació del mètode d'Anàlisi de Components Principals en aquest camp.

Certament, el resultat ha estat positiu pel que fa a la intuïció principal segons la qual aquest mètode hauria de servir també per detectar la presència del xiulet de l'àrbitre en l'anàlisi d'àudio. Malgrat aquest resultat positiu, al llarg del desenvolupament del sistema s'han anat veient possibles millores que no han estat possibles d'implementar per la càrrega temporal que significaven. Aquestes possibles millores s'exposen tot seguit.

Una de les millores a realitzar s'ha vist en el capítol anterior i fa referència a la precisió del sistema. Per tal de millorar la precisió es podria treballar, en primer terme, en una millora del sistema d'entrenament per aconseguir una major qualitat dels xiulets mostra que s'empren en el sistema. El sistema d'entrenament va emprar un conjunt de 52 segments de so de xiulet de diferents partits de futbol però aquests van ser elegits aleatòriament.

Per tal de millorar aquesta característica, es podria pensar en realitzar un sistema que agafés cadascun dels xiulets detectats en l'anàlisi de partits i els incorporés al sistema d'entrenament si trobava que afegien informació rellevant. Concretament es tractaria d'afegir aquells xiulets que al sistema li costen més de detectar (amb un error de reconstrucció més gran).

Una possible segona línia futura de treball contempla la millora del post-processat. Tal i com s'ha pogut veure, l'aplicació del PCA és capaç de detectar la presència dels xiulets de l'àrbitre amb una exhaustivitat del 100% tot i que amb el cost de reduir-ne la precisió. Per això, una millora en el post-processat podria assolir l'objectiu de mantenir l'exhaustivitat al 100% i augmentar alhora la precisió global del sistema.

La millora del post-processat podria realitzar-se a partir de l'anàlisi dels falsos positius per veure com es podrien diferenciar dels veritables positius. Una possibilitat consistiria en fer un segon entrenament del sistema però en sentit invers al primer: entrenar el sistema amb mostres de segments que no corresponen a un xiulet però que han donat un fals positiu per realitzar una segona etapa d'anàlisi amb els positius i contrastar-los de manera que si mostren un error de reconstrucció petit respecte a l'entrenament amb falsos positius aquell segment sigui descartat.

Una altra possibilitat per millorar el post-processat consistiria en adequar millor la detecció dels tres pics de potència que es realitza en la primera etapa del post-processat. El sistema aplica una discriminació i només accepta com a positius aquells segments que mostren tres pics de potència separats per una distància mínima de 100 Hz i amb una potència mínima de 0,2 vegades la

màxima potència del segment estudiat. Si aquesta etapa de processat tingués en compte altres criteris a part dels mencionats segurament es podria millorar l'etapa de post-processat.

Finalment, una última línia futura de treball consistiria en el desenvolupament del sistema per tal que fos capaç de detectar altres models de xiulet a part del Fox 40, com ara el Tornado 2000 de Acme o el model Naysir N1B. Malgrat el Fox 40 sigui el model més emprat amb diferència i el model escollit per moltes competicions europees com el seu xiulet oficial, seria ideal que el sistema fos capaç de detectar diversos models de xiulets.

Aquesta possibilitat es podria aplicar fàcilment en xiulets sense bola donat que normalment es tracta de xiulets dissenyats per produir dos o tres pics de potència en freqüències properes. D'aquesta manera, tan sols realitzant un nou entrenament del sistema a partir de mostres d'àudio del nou xiulet es podria aconseguir que el sistema el detectés.

Un exemple de xiulet sense bola és el xiulet Tornado 2000, xiulet oficial en la NBA (National Basketball Association) i la NFL (National Football League). Així, si es desenvolupés el sistema d'entrenament per aquest xiulet es podria emprar el sistema tan per competicions europees com americanes cobrint així una enorme quantitat d'esdeveniments esportius.

## CONCLUSIONS

Al llarg del present document s'ha mostrat l'aplicació del mètode d'Anàlisi de Components Principals en la detecció del xiulet de l'àrbitre dins el marc de l'extracció automàtica d'esdeveniments d'interès en vídeos esportius, concretament en vídeos de la Lliga Professional de Futbol espanyola.

L'extracció automàtica d'esdeveniments d'interès en vídeos esportius emprà l'anàlisi de vídeo i àudio. En l'anàlisi de l'àudio diferents descriptors són tinguts en compte i també altres elements acústics com és el reconeixement del xiulet de l'àrbitre. Aquest és important perquè acostuma a anar relacionat amb esdeveniments d'interès d'un partit com són l'inici i final del partit, les faltes, els penals, etc., ja que el reglament estableix en quins moments l'àrbitre ha d'emprar el xiulet i en quins casos és recomanable que ho faci.

L'oferta de xiulets en el mercat és àmplia però el model de xiulet oficial de la majoria de competicions europees i el més preferit per la seva fiabilitat és el model Fox 40, un xiulet sense bola que produeix un so compost de tres pics de freqüència.

Donada la importància de conèixer les característiques tècniques del xiulet arbitral per tal d'enfocar correctament la solució al problema s'ha exposat la caracterització realitzada pel xiulet més emprat en la competició d'interès. Aquesta caracterització ha mostrat pistes de camí i alhora reptes a superar. Així, la caracterització ha mostrat que el xiulet Fox 40 presenta tres pics de potència en la banda freqüencial que va dels 3.5 kHz als 4.5 kHz.

Donades aquestes característiques s'han descrit les primeres tècniques emprades per intentar detectar el xiulet Fox 40. Aquests primer mètodes emprats no han aconseguit assolir desitjat degut a una característica del Fox 40 no esperada en un primer moment consistent en la mobilitat en la dimensió freqüencial que presentaven els tres pics de potència que componen el so del xiulet Fox 40. Aquesta mobilitat es presentava en el so d'un mateix xiulet però sobretot en comparar diferents sons de xiulet Fox 40.

Per tal de superar aquesta dificultat s'empra l'Anàlisi de Components Principals, eina emprada per al reconeixement d'objectes complexos. L'objectiu del mètode d'Anàlisi de Components Principals és compactar, amb un error reduït, tota la informació dels vectors en una nova base que utilitza un número reduït de components. D'aquesta manera la nova base filtra el soroll i revela l'estructura comuna. Un cop obtinguda l'estructura fonamental o comuna aquesta es pot aplicar per trobar dades que presentin similitud amb aquesta estructura.

El desenvolupament de la tècnica d'Anàlisi de Components Principals per detectar xiulets de l'àrbitre consisteix en dues etapes i es realitza en Matlab així com també la obtenció dels resultats.

La primera etapa, corresponent a l'entrenament del sistema, consisteix en l'anàlisi de sons de partits de futbol enregistrats prèviament en els quals hi ha presència d'un xiulet de tal manera que el sistema sigui capaç de reconèixer-hi l'estructura comuna o patró fonamental del so del xiulet que es vol detectar. La segona etapa, corresponent a l'anàlisi de la pista d'àudio, se centra en analitzar la pista d'àudio del partit en el qual es volen detectar els xiulets de l'àrbitre.

Les primeres proves mostren que el mètode d'Anàlisi de Components Principals emprat per detectar xiulets d'àrbitre és capaç d'identificar els xiulets de nous partits de futbol a partir de l'estructura comuna obtinguda en l'etapa d'entrenament del sistema a partir dels diferents sons de xiulets enregistrats prèviament.

Tal i com es mostra en el capítol cinquè, el mètode emprat és capaç de detectar els xiulets de l'àrbitre en partits de futbol de la LPF amb una exhaustivitat del 100% però a canvi de reduir la seva precisió a un 12,6%. Això significa que el sistema és capaç de detectar tots els xiulets d'un partit i no produir cap fals negatiu però que en contrapartida augmenta considerablement el nombre de falsos positius.

Els resultats mostren també que una reducció de la precisió al 28,8% significa reduir l'exhaustivitat a 88,2%. Així, un augment de falsos negatius permet reduir els falsos positius.

Per altra banda els resultats també mostren que en ambdós casos, el sistema té una exactitud superior al 99,5%. Això significa que el sistema és capaç d'eliminar el 99,5% dels casos en els quals l'àudio analitzat no presenta la presència d'un xiulet. També s'ha pogut veure en el capítol 5, referent als resultats, que en termes de segments d'àudio analitzats de 4096 mostres, el sistema presenta una precisió propera al 50%.

Per tot això es pot concloure que el mètode d'Anàlisi de Components Principals presenta un bon rendiment quan és aplicat a la detecció de xiulets d'àrbitre en partits de futbol però és discutible si aquest rendiment pot no ser suficient quan es busca no tenir cap fals negatiu i una precisió molt propera al 100%. Per això, per tal d'avaluar si el PCA és una eina vàlida per a la detecció del xiulet de l'àrbitre en l'extracció automàtica d'esdeveniments esportius d'interès, cal sospesar quina precisió i exhaustivitat són necessàries en aquest context.

Cal tenir en compte que el sistema d'extracció d'esdeveniments d'interès empra altres elements d'anàlisi al marge de la detecció del xiulet perquè, tal i com s'ha vist en el capítol primer, no totes les jugades d'interès són causades ni provoquen un xiulet de l'àrbitre. Per tant, la detecció del xiulet de l'àrbitre és una eina que ajuda a que una jugada d'interès sigui detectada però no n'és completament indispensable sinó que altres elements com l'augment de la potència mitja del so o un zoom de la càmera poden ser suficients per detectar l'esdeveniments d'interès.

Per altra banda, la detecció d'un xiulet no necessàriament implica que l'event al qual fa referència hagi de ser seleccionat ja que en més d'una ocasió el xiulet

pot indicar un esdeveniment sense interès, com una falta en el mig del camp o un fora de banda servit incorrectament, etc.

Tot això ens porta a desitjar que en la detecció de xiulets l'exhaustivitat i la precisió siguin el més alts possible però sobretot que mantinguin una relació de compromís adequada amb el conjunt del sistema d'extracció d'esdeveniments esportius d'interès.

Així, i aplicant les millores exposades en el capítol sisè sobre les línies futures de treball, creiem que el PCA es presenta com un mètode vàlid de cara a la detecció de xiulets de l'àrbitre en el context de l'extracció automàtica d'events d'interès en vídeos esportius.

Finalment, notar que l'Anàlisi de Components Principals és una eina emprada en camps diversos però de la qual no en coneixíem aplicació en el camp de la detecció de xiulets d'àrbitre. És per això que, en qualsevol cas, valorem positivament el coneixement del seu comportament aplicat en aquest nou camp, esperant que sigui útil en un futur tant per intentar aconseguir-ne una aplicació més precisa i exhaustiva com per aplicar-ho a altres camps.

## BIBLIOGRAFIA CONSULTADA

- [1] Tarrés, F., Torres, L., Martín, A., "Análisis y resultados con descriptores de audio de bajo nivel". Buscamedia, Programa Cenit-E. Número d'identificació d'expedient CEN-20091026.
- [2] Quijada, R., Raventós, A., Martín, A., Tarrés, F., Torres, L., "Algoritmos de clasificación de eventos simples a partir de descriptores multimodales". Buscamedia, Programa Cenit-E. Número d'identificació d'expedient CEN-20091026
- [3] Kim, H.G., Moreau, N., Sikora, T., *MPEG-7 Audio and Beyond*, John Wiley & Sons, Ltd., Chichester (2005).
- [4] Howard, D.M and Angus, J., *Acoustics and Psychoacoustics*, Focal Press, Oxford (1996).
- [5] <http://www.phys.unsw.edu.au/~jw/graphics/Fletcher-Munson.gif>
- [6] Reglas del Juego 2012/2013, FIFA (Fédération Internationale de Football Association)  
[http://es.fifa.com/mm/document/footballdevelopment/refereeing/81/42/36/lawsofthegame\\_2012\\_s.pdf](http://es.fifa.com/mm/document/footballdevelopment/refereeing/81/42/36/lawsofthegame_2012_s.pdf)
- [7] <http://www.enotes.com/whistle-66422-reference/whistle>
- [8] [http://www.fox40world.com/index.cfm?pagepath=ABOUT\\_US/Who\\_We\\_are\\_\\_The\\_Fox\\_40\\_Story&id=4099](http://www.fox40world.com/index.cfm?pagepath=ABOUT_US/Who_We_are__The_Fox_40_Story&id=4099)
- [9] [http://www.Fox40world.com/index.cfm?DSP=ProductDetail&ProductID=126&SubCatID=0&CategoryID=1002&pagepath=PRODUCTS/Product\\_Catalog&id=4130](http://www.Fox40world.com/index.cfm?DSP=ProductDetail&ProductID=126&SubCatID=0&CategoryID=1002&pagepath=PRODUCTS/Product_Catalog&id=4130).
- [10] Proakis, J.G., Manolakis, D.G., *Advanced digital signal processing*, Prentics-Hall International, Inc., Upper Saddle River, New Jersey, (1995).
- [11] Bishop, C. M., *Pattern recognition and machine learning*, Springer Science and Business Media, LLC, New York (2006).
- [12] Duda, R. O., Hart, P.E., Stork, D.G., *Pattern classification*, John Wiley & Sons, Inc, New York (2001).
- [13] MacKay., D.J.C., *Information Theory, Inference, and learning Algorithms*, Cambridge University Press, Cambridge (2003).

- [14] Oppenheim, A.V. , Schaefer, R.W., i Buck, J.R., *Digital signal processing*, Prentice-Hall, Inc., Upper Saddle River, New Jersey (1999).
- [15] <http://publib.boulder.ibm.com/infocenter/classify/v8r7/index.jsp?topic=%2Fcom.ibm.classify.common.doc%2Fbnrgl000.htm>
- [16] ESS Handbook for Quality Reports, European Communities, 2009.  
[http://147.67.136.2/f9c2d23df842/comfortfile/75de\\_4fe0\\_0e108450\\_5b41\\_11\\_e2\\_bb93\\_001b21a4d300/MTcyLjE2LjEuMw==/EHQR\\_FINAL.pdf](http://147.67.136.2/f9c2d23df842/comfortfile/75de_4fe0_0e108450_5b41_11_e2_bb93_001b21a4d300/MTcyLjE2LjEuMw==/EHQR_FINAL.pdf)
- [17] <http://publib.boulder.ibm.com/infocenter/classify/v8r7/index.jsp?topic=%2Fcom.ibm.classify.common.doc%2Fbnrgl000.htm>